



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

CAMPUS D'ALCOI

# UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Politécnica Superior de Alcoy

Modelización Econométrica de los Factores Determinantes  
de la Calidad del Vino

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Serra Izquierdo, Araceli

Tutor/a: Hervás Marín, David

Cotutor/a: Carracedo Garnateo, Patricia

CURSO ACADÉMICO: 2023/2024

# Resumen

El vino es y ha sido a lo largo de la historia una bebida popular en las diferentes sociedades. El proceso de producción del vino se ha ido refinando con el paso de los años hasta el punto de llegar a estudiar qué componentes añadir y la cantidad exacta del mismo para conformar la matriz del vino idónea que haga ser percibido al vino como de los mejores para el consumidor. Por ello, las empresas especializadas en vinos que han ido surgiendo a lo largo del tiempo tienen muy presente la matriz del vino que eligen y las opiniones de los usuarios sobre la misma.

El objetivo de este trabajo fin de grado es estudiar cómo la matriz del vino y las opiniones de expertos en catas sobre los componentes del mismo afectan a su calidad final percibida y, por tanto, permitirle a la empresa conocer la mejor forma de corregir el producto para que sea más agradable para el consumidor y con ello poder aumentar las ventas de su producto. Para alcanzar este objetivo se han aplicado técnicas econométricas y de machine learning basadas en la regresión ordinal. En este caso, se ha usado R como software estadístico, puesto que es un lenguaje especializado en el análisis, tratamiento y visualización de datos.

Palabras clave: vino, R, regresión ordinal, machine learning, análisis de datos, cata, calidad, componentes, matriz del vino.



# Resum

El vi és i ha sigut al llarg de la història una beguda popular en les diferents societats. El procés de producció del vi s'ha anat refinant amb el pas dels anys fins al punt d'arribar a estudiar quins components afegir i la quantitat exacta del mateix per a conformar la matriu del vi idònia que faça ser percebut al vi com dels millors per al consumidor. Per això, les empreses especialitzades en vins que han anat sorgint al llarg del temps tenen molt present la matriu del vi que trien i les opinions dels usuaris sobre esta.

L'objectiu d'este treball fi de grau és estudiar com la matriu del vi i les opinions d'experts en tastos sobre els components del mateix afecten la seua qualitat final percebuda i, per tant, permetre-li a l'empresa conèixer la millor manera de corregir el producte perquè siga més agradable per al consumidor i amb això poder augmentar les vendes del seu producte. Per a aconseguir este objectiu s'han aplicat tècniques econòmiques i de *machine learning* basades en la regressió ordinal. En este cas, s'ha usat R com a programari estadístic, ja que és un llenguatge especialitzat en l'anàlisi, tractament i visualització de dades.

Paraules clau: vi, R, regressió ordinal, *machine learning*, anàlisi de dades, tast, qualitat, components, matriu del vi.



# Abstract

Wine is and has been throughout history a popular drink in different societies. The wine production process has been refined over the years to the point of studying what components to add and the exact amount thereof to form the ideal wine matrix that makes the wine perceived as one of the best for the consumer. For this reason, the companies specialized in wines that have emerged over time are very aware of the matrix of the wine they choose and the opinions of users about it.

The objective of this final degree project is to study how the matrix of the wine and the opinions of experts in tastings about its components affect its final perceived quality and, therefore, allow the company to know the best way to correct the product. to make it more pleasant for the consumer and thereby increase the sales of your product. To achieve this objective, econometric and machine learning techniques based on ordinal regression have been applied. In this case, R has been used as statistical software, since it is a language specialized in data analysis, processing and visualization.

Keywords: wine, R, ordinal regression, machine learning, data analysis, tasting, quality, components, wine matrix.



# Índice de figuras

1.1. Desagregación del impacto del sector en el empleo, por tipo de efecto (empleos). <b>Fuente:</b> Internacionales 2023 . . . . .	2
1.2. Producción media de uva para vinificación, por provincia. <b>Fuente:</b> Internacionales 2023 . . . . .	3
1.3. Crecimiento demográfico municipal, por perfil de municipio, según tamaño e intensidad de viñedo, reflejando la variación de los años 2002-2022. <b>Fuente:</b> Internacionales 2023 . . . . .	3
1.4. Principales estructuras químicas y origen de los taninos enológicos. <b>Fuente:</b> Versari, Du Toit y Parpinello 2013 . . . . .	5
1.5. Mecanismo de percepción del sabor en la ingesta de alimentos y vinos. <b>Fuente:</b> Redondo, Gómez-Martínez y Marcos 2014 . . . . .	7
1.6. Rueda de los aromas del vino utilizada por catadores. <b>Fuente:</b> <i>La rueda de los aromas del vino</i> s.f. . . . .	8
1.7. Variación del color del vino según la cantidad de oxígeno presente. <b>Fuente:</b> Jacqui M McRae, Day et al. 2015 . . . . .	9
1.8. Recuento celular total en cámara de Neubauer. Imagen de microscopía óptica, x400 aumentos. A- Barrica sin tratar, contaminación inicial superior a 107 UF-C/ml. B- Barrica anterior tratada. <b>Fuente:</b> Palacios y Beaucourt s.f. . . . .	10
3.1. Correlación entre las variables calidad y astringencia. . . . .	21
3.2. Correlación entre la calidad y el alcohol percibido. . . . .	21
3.3. Correlación entre la calidad y la varibale calidad del aroma percibida. . . . .	21
3.4. Correlación entre la calidad y la planta (porte) del vino. . . . .	22



3.5. Correlación entre la calidad y la intensidad visual percibida del vino. . . . .	22
3.6. Correlación entre la variable calidad y sabrosidad del vino percibidos. . . . .	22
3.7. Correlación entre el alcohol y la intensidad del aroma percibidos. . . . .	23
3.8. Correlación entre la intensidad del olor percibido y la intensidad aroma percibida.	23
3.9. Correlación entre la variable densidad del vino y alcohol que contiene el mismo.	23
3.10. Matriz de correlación para correlaciones superiores a 0.4, según el algoritmo de diseño Fruchterman-Reingold. . . . .	24
3.11. Matriz de correlación para correlaciones superiores a 0.8, según el algoritmo de diseño Fruchterman-Reingold. . . . .	25
3.12. Gráfica de efectos para la variable <i>chlorides</i> . . . . .	28
3.13. Gráfica de efectos para la variable <i>Phenolic</i> . . . . .	28
3.14. Gráfica de efectos para la variable <i>density</i> . . . . .	29
3.15. Gráfica de efectos para la variable <i>Aroma.persistency</i> . . . . .	29
3.16. Gráfica de efectos para la variable <i>Acidity</i> . . . . .	29
3.17. Gráfica de efectos para la variable <i>Aroma.quality</i> . . . . .	30
3.18. Gráfica de efectos para la variable <i>Attack</i> . . . . .	30
3.19. Gráfica de efectos para la variable <i>Balance</i> . . . . .	31
3.20. Gráfica de efectos para la variable <i>Plante</i> . . . . .	32
3.21. Gráfica de efectos para la variable <i>Smooth</i> . . . . .	32

# Índice de tablas

3.1. Análisis descriptivo de la base de datos. . . . .	20
3.2. Resultados obtenidos tras ejecutar el comando <i>summary(model1)</i> . . . . .	26
3.3. Resultados obtenidos tras ejecutar el comando <i>summary(model2)</i> . . . . .	26
3.4. Resultados obtenidos tras ejecutar el comando <i>summary(model3)</i> . . . . .	26
3.5. Resultados obtenidos tras ejecutar el comando <i>summary(model4)</i> . . . . .	27
3.6. Resultados obtenidos tras ejecutar el comando <i>summary(model5)</i> . . . . .	27
3.7. Resultados obtenidos tras ejecutar el modelo con la corrección de Lipsitz. . . . .	28



# Índice general

Resumen	I
Índice general	XI
1 Introducción	1
1.1 Importancia histórica . . . . .	1
1.2 Importancia del sector vitivinícola . . . . .	2
1.3 Componentes. . . . .	3
1.4 Importancia empresarial (Motivación y objetivos) . . . . .	10
2 Metodología	11
2.1 Obtención de la información . . . . .	11
2.2 Modelo de regresión lineal . . . . .	15
2.3 Regresión ordinal . . . . .	16
3 Resultados	19
3.1 Análisis Descriptivo . . . . .	19
3.2 Desarrollo del modelo . . . . .	25
3.3 Interpretación . . . . .	28
4 Conclusiones	33
Bibliografía	37



## Capítulo 1

# Introducción

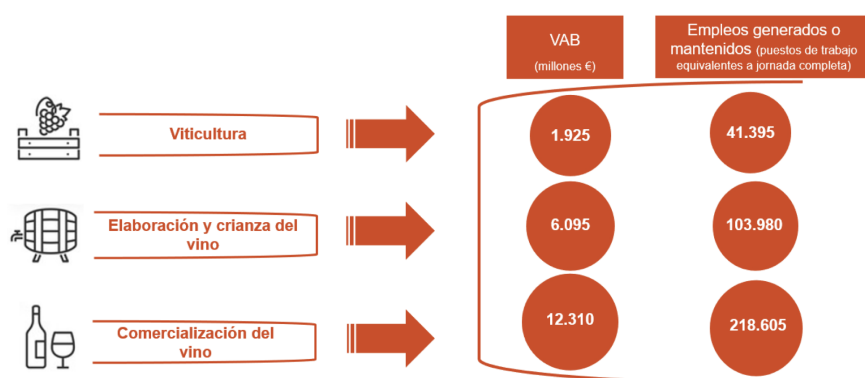
### 1.1 Importancia histórica

Históricamente hablando, el vino siempre ha sido una bebida icónica para el ser humano puesto que, desde el año 8.000 a.C, pasando por la antigua Grecia y Roma y hasta la actualidad (NEWS s.f.), el hombre ha consumido vino en muchos momentos de su vida, tanto positivos como negativos e independientemente del acontecimiento que sucediese en el día a día, como se plasma en las obras de arte de *La última cena* de Leonardo da Vinci o en *Los borrachos* de Velázquez. Este líquido ha sido tan importante que incluso en aquella época, los griegos adoraban a Dionisio, su dios del vino, al igual que los romanos celebraban la festividad en honor a Baco, el que era también el dios del vino (Fehér, Lengyel y Lugasi 2007). Independientemente de la cultura y el siglo en el que se viviese, el vino ha sido un elemento imprescindible de la sociedad, y la tradición se ha mantenido así durante siglos hasta llegar a la actualidad. En cambio, el proceso de fabricación del vino sí ha ido perfeccionándose con el paso de los años, evolucionando desde el prensado o pisado de uva realizado por los monjes con sus pies en la edad media, los cuales, posteriormente almacenaban y fermentaban el vino de forma natural en las primeras barricas, sin poder evitar que con la llegada de la primavera el vino se convirtiese en vinagre (R. S. Jackson 2008); hasta que el proceso de fabricación del vino ha pasado a ser considerado arte y ciencia basado en medidas y correcciones con la tecnología más puntera para la obtención del mejor vino del mercado.

La experiencia de consumir un buen vino no se centra solo en el momento en el que se bebe este, si no que comienza en la selección y apertura de la botella, pasando por la percepción de todos los componentes que forman la matriz del vino, es decir aromas, textura, color y densidad, todo ello captado de forma olfativa y visual, y termina con los sabores y textura percibidos a través del gusto. Cada detalle está estudiado y medido para que sea perfecto y guste al mayor número de consumidores, sin considerar que cada persona tiene un gusto diferente y prefiere consumir un tipo de vino distinto a otro cliente.

## 1.2 Importancia del sector vitivinícola

En España, el sector vitivinícola es considerado como relevante por motivos culturales, sociales y económicos. Según los estudios sobre la relevancia económica del sector vitivinícola promovidos y publicados por la Interprofesional del Vino de España (OIVE), en el año 2023 el sector contribuyó en la generación de más de 20.330 millones de euros, lo que supone aproximadamente un 2 % del PIB nacional a precios de mercado. Para que esto sea posible, el sector genera cerca de 363.980 empleos a tiempo completo, lo que se traduce en un 2 % del empleo nacional (Internacionales 2023). Ambas cifras se pueden observar en la figura 1.1. La creación de tantos puestos de trabajo en este sector es beneficioso para el país porque, entre las características de estos empleos, destaca la dedicación a largo plazo que se ha de desempeñar, generando así que la población se asiente durante un periodo largo de tiempo en el mismo territorio (Sánchez-Hernández 2014, Internacionales 2023).



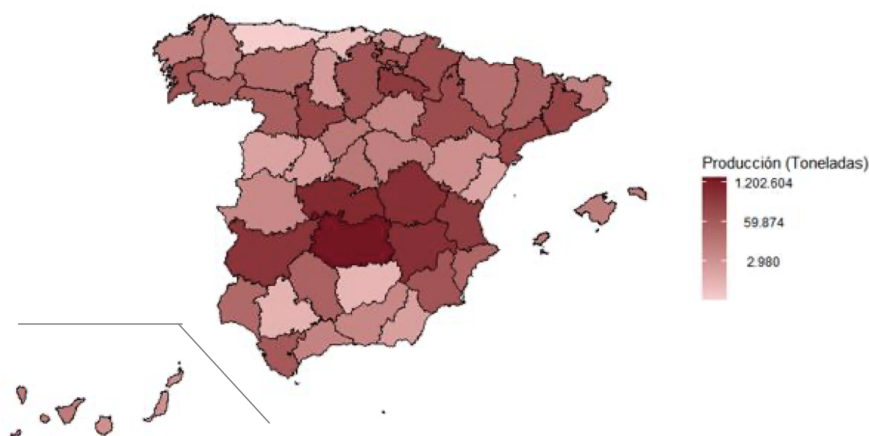
**Figura 1.1:** Desagregación del impacto del sector en el empleo, por tipo de efecto (empleos).

**Fuente:** Internacionales 2023

Por otro lado, todas las comunidades autónomas de España cuentan con una mayor o menor cantidad de terrenos de cultivo vitivinícola (figura 1.2), por lo que se cuenta con un total de más de 930.000 hectáreas de viñedos en el país caracterizadas por dividirse en explotaciones pequeñas y rurales, de las cuales 142.100 son viñedos ecológicos, situando a España como uno de los países líderes en el sector. Estos datos geográficos representan a nivel mundial un 13 % de las hectáreas destinadas a este cultivo (Internacionales 2023).

Esta cantidad de cultivos producen de forma anual un montante de 36.4 millones de hectolitros de vino caracterizado por su calidad y diferenciación de la competencia. Y no solo se sitúa a España en esa posición de liderazgo frente a países como Francia e Italia por la cantidad de hectáreas que posee (Portela 2013), sino también por el gran valor añadido percibido en el vino por los clientes extranjeros (De la Torre Pigazos 2019) y por el volumen de exportación de producto con denominación de origen que se realiza al extranjero (2.153 millones de toneladas en 2022), que generaron en los últimos años un superávit comercial por valor de 3.058 millones de euros (Internacionales 2023).

Además de ser el motor de una parte significativa de los ingresos en España, el sector vitivinícola también promueve un gran número de empleos, como bien ya se ha mencionado anteriormente, en las localidades más rurales, lo que desencadena que, las zonas rurales que cuentan con cultivo y producción de vid están volviendo a incrementar su población en un 36 % en comparación con las cifras de hace 20 años, ayudando de esta forma a que la tendencia creciente del éxodo rural soportado por las poblaciones rurales durante estos años se reduzca o incluso se revierta en las localidades con esta característica (Internacionales 2023), (Sánchez-Hernández 2014). Más concretamente, se observa en la figura 1.3 que durante el año 2002 hasta el 2022, se produjo



**Figura 1.2:** Producción media de uva para vinificación, por provincia.

**Fuente:** Internacionales 2023

una disminución de aproximadamente el 10 % en pequeñas poblaciones sin cultivos de viñedos, mientras que aquellas poblaciones clasificadas en el mismo rango de tamaño que sí tenían este tipo de terreno, su número de habitantes aumentó en un 11,4 % (Internacionales 2023).

Tamaño poblacional / Intensidad de viñedo	Menos de 2.000 hab	Entre 2.000 y 5.000 hab	Entre 5.000 y 30.000 hab	Más de 30.000 hab	Total
Sin viñedo y con presencia compartida de viñedo con otros cultivos	-10,2%	9,1%	22,6%	12,8%	13,3%
Con viñedo como cultivo principal (>50% de la superficie agrícola)	11,4%	20,3%	24,6%	19,3%	21,2%
<b>Total</b>	<b>-9,9%</b>	<b>9,4%</b>	<b>22,6%</b>	<b>12,8%</b>	<b>13,4%</b>

**Figura 1.3:** Crecimiento demográfico municipal, por perfil de municipio, según tamaño e intensidad de viñedo, reflejando la variación de los años 2002-2022.

**Fuente:** Internacionales 2023

## 1.3 Componentes

Para que el vino sea percibido como de máxima calidad, todos los aspectos sensoriales mencionados anteriormente han de estar en un perfecto estado de equilibrio, y para ello se han de tener en cuenta muchas variables que intervienen en el proceso de fabricación. La matriz de vino anteriormente mencionada es el conjunto de componentes que conforman un vino, diferenciados en matriz volátil y no volátil. La matriz volátil está formada por compuestos de sabor y aroma, mientras que la fracción no volátil, por etanol (en fase líquida), compuestos polifenólicos, proteínas e hidratos de carbono (Villamor y Ross 2013).



### 1.3.1 Taninos

Uno de estos componentes que es muy influyente en diferentes integrantes de la matriz del vino es la cantidad de taninos existentes en el mosto o vino.

Los taninos son compuestos polifenólicos que se encuentran en la piel, semillas y tallo de la uva, además de aparecer en la madera de roble con la que se fabrican algunas de las barricas utilizadas en el proceso de envejecimiento del vino (Castaldo et al. 2019).

Son los causantes de ocasionar al vino tinto la sensación de astringencia y amargura, definir el color que tendrá y además son un antioxidante natural que permite que el producto se preserve y envejezca adecuadamente (Lachman et al. 2009).

Los efectos antioxidantes no solo aportan beneficios al vino, pues en el ser humano ayudan a reducir las enfermedades cardiovasculares, a causa de la disminución del colesterol LDL al reducir su oxidación (Castaldo et al. 2019), la agregación de plaquetas, la osteoporosis, también actúan como agentes antiinflamatorios y moduladores de rutas metabólicas (Suárez et al. 2018), en enfermedades como la artritis y diabetes respectivamente, todo ello siempre y cuando se tome una cantidad de vino moderada.

La astringencia provocada por los taninos del vino se define como la sensación de sequedad, aspereza y rugosidad de las mucosas bucales, las cuales son ocasionadas por la precipitación de la proteína salivar cuando se produce el enlace tanino-proteína salivar, pudiendo causar modificaciones de la percepción del sabor del vino y de los alimentos proteicos que se ingieren con este (Jacqui M McRae y Kennedy 2011).

La presencia de taninos es notablemente mayor en los vinos tintos, mientras que en los vinos blancos la concentración de estos es significativamente más baja. Una de las razones de que suceda esto es el tiempo que dura la etapa de maceración. La duración de esta etapa en los vinos blancos es muy corta o inexistente, impidiendo así la adherencia de los taninos al mosto por la rápida separación que se lleva a cabo de este con las pieles y pepitas de la uva. Esto también afecta en el desarrollo de las etapas posteriores de fermentación y envejecimiento, porque la primera se lleva a cabo sin las semillas y pieles de la uva, y el tiempo que pasa el vino blanco en barricas de roble en la etapa de envejecimiento es menor que el que pasa el vino tinto, causando que los taninos existentes en las barricas no se lleguen a transferir.

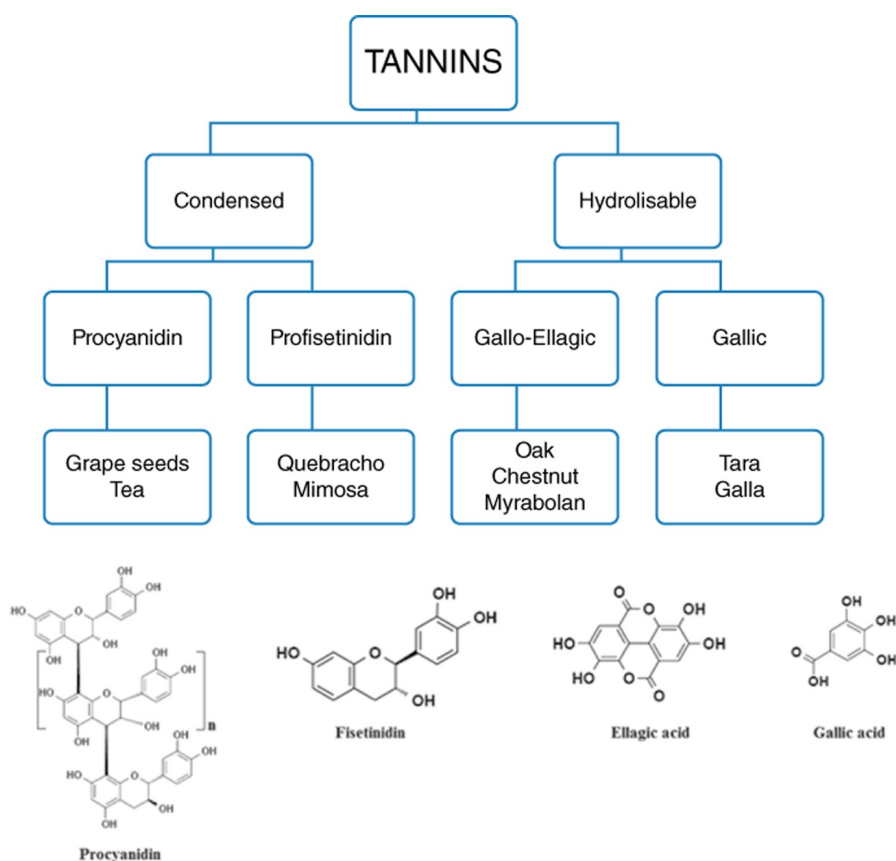
Otra de las causas que provoca esta diferencia de concentración de taninos es la influencia que tiene la variedad de uva a utilizar para la fabricación de cada vino, puesto que variedades como el Chardonnay para vinos blancos tienen una menor concentración de taninos que las usadas para la elaboración de vinos tintos.

Una vez se ha definido el concepto de tanino, se ha de matizar que existen dos tipos de los mismos (figura 1.4). Los hidrolizables son taninos con naturaleza química de éster de ácido fenólico, caracterizados por ser más pequeños molecularmente además de más amargos y astringentes que el otro tipo, los taninos condensados. Estos aparecen cuando el vino está envejeciendo en las barricas de roble, aumentando la cantidad existente a medida que el vino reposa durante más tiempo en estos recipientes además de tener en cuenta el desgaste de las barricas y el origen del roble con el que está fabricado (Smith, J. McRae y Bindon 2015).

Por otro lado, los taninos condensados, también conocidos como taninos de uva, son proantocianidinas de tamaño molecular más grande que los hidrolizables. Los taninos de uva provienen de la piel, la carne y las semillas de la uva y son extraídos en el momento de elaboración del

vino. Contribuyen hasta 4 g/L al vino tinto con la mayoría de los taninos (Smith, J. McRae y Bindon 2015), además de aportar al vino un sabor menos astringente y amargo.

Estos últimos tienen una variante conocida como taninos de vino, los cuales se diferencian por ser taninos modificados mediante las correcciones de levaduras y enzimas aplicadas al mosto durante el proceso de fermentación. Con el paso del tiempo, los taninos del vino se vuelven menos astringentes, dado que continúan experimentando cambios químicos que también acaban modificando de forma gradual el tono púrpura del vino joven a un color rojo ladrillo (Smith, J. McRae y Bindon 2015).



**Figura 1.4:** Principales estructuras químicas y origen de los taninos enológicos.

**Fuente:** Versari, Du Toit y Parpinello 2013

Dependiendo de la procedencia de los taninos, se pueden clasificar como taninos naturales y comerciales, independientemente de si son condensados o hidrolizables.

Los taninos naturales proceden de medios naturales como la uva o cualquier especie vegetal con especial concentración en árboles como el roble o el quebracho. En el vino aparecen mayoritariamente los taninos de uva y es imprescindible saber controlar los tiempos en los que el vino está en contacto con estos componentes para que aporten el nivel de propiedades justas al producto. Los comerciales son sintéticos y procesados de una forma refinada y purificada, tras la extracción forzada de medios naturales como el quebracho y el roble (Versari, Du Toit y Parpinello 2013), para la venta destinada a muchas industrias diferentes. Estos taninos comerciales son utilizados para realizar las correcciones deseadas a los vinos, dado que sus efectos sobre este líquido son mucho más predecibles y controlables que si fuesen naturales, pero en cambio, su efecto es mayor que el de los taninos naturales y por tanto la astringencia captada también lo es.

Es importante conocer que los taninos pueden tener una cierta calidad que afecta al grado de astringencia y aromatización, al igual que a la intensidad percibida de los sabores afrutados y el cuerpo del vino. A nivel general, se considera que un tanino es de mayor calidad cuando este es suave y sedoso, mientras que la calidad es menor si son ásperos y astringentes. La variedad de uva que se utilice en la producción del vino es una de las principales causas de esta diferencia de calidad, como sucede con las variedades de uvas Cabernet Sauvignon y Nebbiolo, que se diferencian de la variedad Pinot Noir por tener una piel más gruesa y por tanto más rica en taninos.

Otro motivo que determina la calidad tánica es el nivel de madurez de la uva cuando se utiliza para la elaboración del vino puesto que, si se cosecha de forma temprana va a tener más taninos y producirá un vino más astringente, en cambio si se cosecha la uva más madura, se obtendrán vinos más suaves. Para una maduración completa de la uva también influye el clima en el que se cultiva ya que, a más cálido su cultivo y maduración será más completo (Lachman et al. 2009).

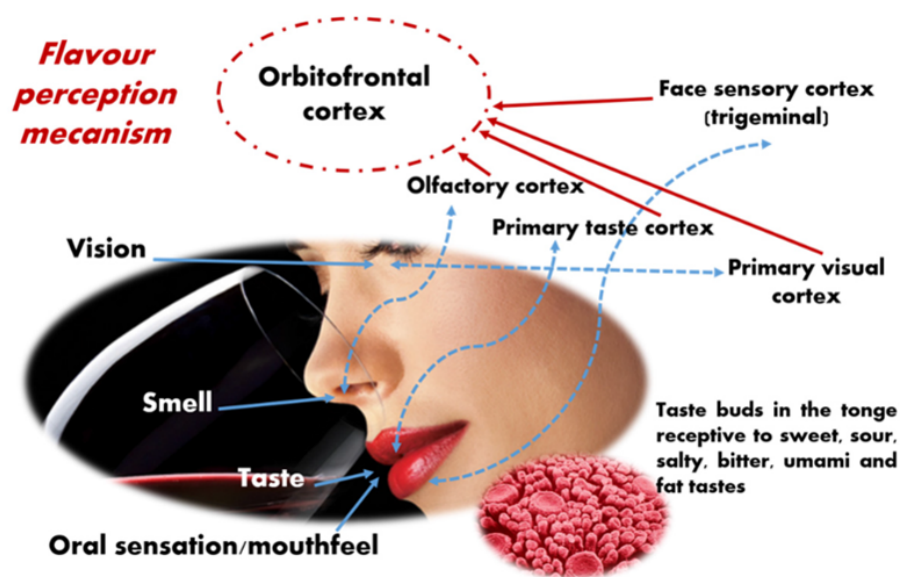
### **1.3.2 Aromas y alcohol**

La aromatización del vino es otro de los aspectos que se tiene en cuenta para definir la calidad de un vino, y es de los factores más decisivos a la hora de que un consumidor escoja el vino que va a tomar (Intensos y Richardi s.f.), además dependiendo de los aromas que un vino proporcione, su precio establecido aumenta, como pasa con la presencia de aromas a madera (Juan et al. 2012). Esta puede ser modificada o eclipsada por los taninos, como se ha mencionado anteriormente, al igual que los microorganismos de la uva y otros añadidos también afectan a la aromatización (Carpena et al. 2020). Cabe destacar que, en el mundo de la enología, existe diferencia entre los términos “aroma” y “olor”, puesto que un conjunto de olores percibidos, refinados e identificados conforman un aroma (Ferreira, De-La-Fuente-Blanco y Sáenz-Navajas 2021).

En la enología, se diferencian varios tipos de aromas, los conocidos como aromas primarios, que provienen de las diferentes variedades de las uvas y al igual que los taninos, dependiendo de factores como el clima, tipo de suelo y maduración variarán los aromas que aporta una misma variedad de uva al vino (G Pereira et al. 2020), (Juan et al. 2012), además de que esta última también afectará a la concentración final de alcohol (Muñoz et al. 2006). Por otro lado, están los aromas secundarios provenientes de las reacciones químicas que se producen en el proceso de fermentación y que por tanto van a depender de la levadura encargada de liberar los aromas que se utilice en dicho proceso. Por último, los aromas terciarios aparecen en el proceso de envejecimiento ya que, al almacenarse el producto durante bastante tiempo en barricas, el vino adquiere los aromas que desprende la madera con la que están fabricadas (Carpena et al. 2020).

Al ser un compuesto volátil, tal y como se ve en la figura 1.5, “los aromas son transportados durante la respiración a través de las fosas nasales, donde interactúan con los receptores olfativos (vía orthonasal). Sin embargo, durante el consumo, los compuestos volátiles se liberan en el interior de la cavidad oral y llegan, gracias a los flujos respiratorios contracorriente a los receptores olfativos (vía retronasal). Esta diferencia de ruta de transporte de los compuestos volátiles hace que la percepción del aroma al oler o al consumir el vino, pueda ser muy diferente” (Intensos y Richardi s.f.).

En el proceso de fermentación alcohólica las levaduras existentes en el mosto o las añadidas por el enólogo son las encargadas de transformar los azúcares provenientes de la uva y existentes en el mosto, en alcohol principalmente y por ende, el estado de madurez de la uva determina la cantidad de alcohol que contiene el vino (Jordão, Vilela y Cosme 2015).



**Figura 1.5:** Mecanismo de percepción del sabor en la ingesta de alimentos y vinos.

**Fuente:** Redondo, Gómez-Martínez y Marcos 2014

Durante este proceso también se desarrollan parte de los aromas del vino, como los aromas secundarios mencionados anteriormente. El etanol que se ha liberado se encarga de fijar el olor, equilibrar el sabor y aportar viscosidad al vino (al igual que lo hacen los taninos), mientras que el glicerol se encarga de la síntesis de los aromas ya mencionados, por lo que una mayor concentración de alcohol en el vino puede modificar la percepción de los aromas así como el sabor del vino por los estímulos percibidos por el paladar (Carpena et al. 2020), (Jordão, Vilela y Cosme 2015).

También se ha de destacar que, “la mayor concentración de alcohol juega un papel clave en la complejidad de la composición del aroma. Los valores óptimos de alcohol por debajo de 300 mg/L proporcionan notas afrutadas y floridas, mientras que los valores de alcohol superiores a 400 mg/L se vuelven negativos al añadir aromas picantes y desagradables” (Perestrelo et al. 2020), (Jeromel, Korenika y Tomaz 2019), (Belda et al. 2017). Esta cantidad alcohólica además de determinar la composición del aroma, también determina la densidad del vino, puesto que, como el alcohol tiene una densidad menor a la del agua, ante una mayor graduación alcohólica, la densidad del vino es menor (Martín 2017).

Además de identificar los diferentes aromas que componen un vino, en la enología también se tienen en cuenta otras características como la calidad, intensidad y persistencia del aroma. Esta última característica mencionada puede ser valorada en varios momentos de la degustación de un vino como por ejemplo, de forma olfativa cuando se descorcha la botella o de forma degustativa cuando se consume el producto.

El alcohol, también conocido científicamente como etanol, junto con otros polifenoles, es un componente del vino que afecta a la retención del aroma de forma oral, lo que deriva en la percepción de la persistencia del aroma.

“El vino con etanol moderado y bajo contenido de polifenol y el vino con bajo contenido de etanol y polifenol moderado presentaron una mayor persistencia de éster oral con respecto al vino de control (bajo contenido de etanol/polifenol). Sin embargo, lo más probable es que el vino con un contenido moderado de etanol/polifenoles muestre una menor persistencia del éster oral en comparación con el resto de los vinos. Por lo tanto, una interacción entre el etanol

y los polifenoles a concentraciones específicas podría reducir la persistencia del éster oral y, probablemente, el carácter afrutado de los vinos” (Muñoz-González, Pérez-Jiménez y María Ángeles Pozo-Bayón 2020).

Lo mismo sucede con la intensidad del aroma percibida de forma oral, pues si el vino a consumir contiene una mayor cantidad de etanol, la intensidad con la que el consumidor percibe el aroma será menor y por tanto, la persistencia del mismo también lo será (Perez-Jiménez, Chaya y Maria Ángeles Pozo-Bayón 2019).

Los aromas escogidos para caracterizar un vino son tan importantes y valorados por el consumidor que existe una rueda de aromas del vino utilizada tanto por los catadores profesionales como por los amateurs para poder identificar con precisión los diferentes aromas presentes en el vino (figura 1.6). Esta rueda de aromas ayuda a identificar primero la familia principal a la que pertenece el aroma y luego a profundizar en cada uno durante el proceso de degustación, hilando cada vez más fino en la percepción de este (*La rueda de los aromas del vino* s.f.).



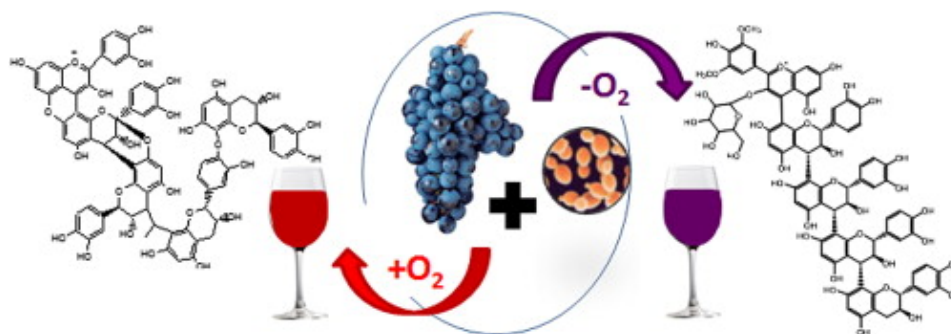
Figura 1.6: Rueda de los aromas del vino utilizada por catadores.  
Fuente: *La rueda de los aromas del vino* s.f.

### 1.3.3 Color

Otro efecto que tienen las levaduras utilizadas en el proceso de la fermentación alcohólica es la modificación y estabilización del color del vino (Carpena et al. 2020 y Padilla, Gil y Manzanares 2016). Para constituir el equilibrio perfecto en esta variable, no influyen solo las levaduras anteriormente mencionadas, si no que, los taninos, el nivel de madurez de la uva y la oxigenación del vino también contribuyen a la especificación del color.

Si la uva se cosecha en una etapa temprana de su madurez, esta no ha podido desarrollar todos los azúcares que tiene una uva madura, lo que desencadena que la intensidad del color sea menor por su menor concentración de pigmentos en la materia prima, al igual que sucede con los aromas (Jordão, Vilela y Cosme 2015).

Por otro lado, la diferente exposición de oxígeno durante el proceso de fermentación de un vino modifica el color final del mismo, puesto que, si se expone durante un plazo de tiempo más prolongado y con cantidades de oxígeno menores y más controladas se obtienen vinos de color rojizo, además de sensaciones bucales más suaves y refinadas, tal y como se puede observar en la figura 1.7. En cambio, si se expone el líquido a grandes cantidades de oxígeno durante un corto periodo de tiempo, el color obtenido es más amoratado (Jacqui M McRae, Day et al. 2015).



**Figura 1.7:** Variación del color del vino según la cantidad de oxígeno presente.

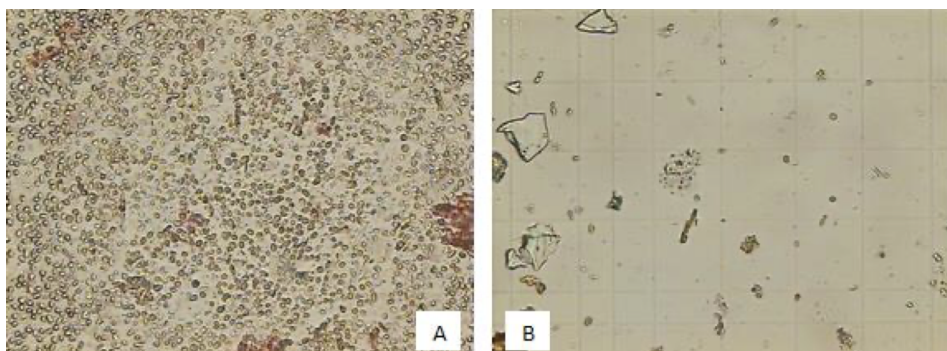
**Fuente:** Jacqui M McRae, Day et al. 2015

Tal y como se ha mencionado anteriormente, los taninos también influyen en esto, ya que en una situación como la que se explica en el primer caso, la concentración de estos disminuye ocasionando una deficiencia en la intensidad del color totalmente perceptible para el ojo humano (Alcalde-Eon et al. 2014). Por otro lado, la adición de taninos provenientes del quebracho a los proporcionados por la misma uva utilizada para la fabricación del vino, ocasiona que la intensidad del color se modifique positivamente (Hernández Orte s.f.). Añadido a esto, la temperatura a la que se consume el vino también es un factor a tener en cuenta para determinar el color del vino ya que, a temperaturas frías, los taninos se vuelven menos perceptibles y los pigmentos se condensan haciendo que el color sea menos intenso (Otamendi s.f.).

## 1.4 Importancia empresarial (Motivación y objetivos)

Tal y como se ha explicado anteriormente, el vino es un producto que asume un papel importante en la sociedad, independientemente del país en el que se resida, y es por este motivo que las empresas fabricantes, con el paso del tiempo han ido y siguen perfeccionando sus técnicas de elaboración para poder ofrecer el mejor producto con la mejor calidad al consumidor. Para conseguir este objetivo, las empresas vinícolas destinan una gran inversión al departamento de I+D (Martínez Montero et al. 2020), pero no solo es importante destinar una gran cantidad de dinero, si no que también es imprescindible conocer las propiedades de los diferentes componentes del vino, cómo se comportan a lo largo del proceso de fabricación y cómo son percibidos al consumir el vino, para que de este modo, la empresa fabricante pueda corregir de una forma determinada los componentes de la matriz del vino que desencadenan las características del mismo, para que el producto obtenga una mejor calidad.

Por ejemplo, una de las innovaciones que se está investigando y llevando a cabo, es un nuevo diseño de tapones de botella que dejan pasar el oxígeno para que la tanosidad disminuya pero se mantenga la percepción del color (Gambutí et al. 2017). Otra innovación es la regeneración de las barricas con 3 o 4 años de uso mediante un método llamado Barena, para conseguir reactivar las funciones enológicas que estos depósitos tienen y han perdido con el paso del tiempo a causa del uso (figura 1.8). De esta forma se permite alargar la vida útil de estos activos en las empresas vitícolas y seguir aportando al vino las características tánicas y aromáticas de la misma forma que lo hace una barrica nueva (Palacios y Beaucourt s.f.).



**Figura 1.8:** Recuento celular total en cámara de Neubauer. Imagen de microscopía óptica, x400 aumentos. A- Barrica sin tratar, contaminación inicial superior a 107 UFC/ml. B- Barrica anterior tratada.

**Fuente:** Palacios y Beaucourt s.f.

El objetivo de este Trabajo Fin de Grado es estudiar mediante un análisis estadístico, la relación que existe entre las características percibidas por catadores profesionales en la cata de diferentes vinos, con la puntuación final de la calidad que asignan a cada uno. Esto resulta interesante para las empresas vitivinícolas ya que, pueden conocer mejor la interacción de los componentes del vino y cómo esta se correlaciona con la calidad final percibida por los consumidores. Con ello, se les permitirá mejorar y optimizar sus procesos de producción y ofrecer productos de mayor calidad, alineados con las expectativas y preferencias de los consumidores.

Este estudio también es interesante para las empresas emergentes en el sector vinícola, ya que van a poder tener conocimientos sobre las ya mencionadas relaciones entre los componentes y la calidad para elaborar un producto que se adapte mejor a los requisitos del consumidor desde la creación y lanzamiento del mismo, generando entre los consumidores una buena imagen y reputación de marca.

# Metodología

### 2.1 Obtención de la información

Se ha realizado un análisis y estudio estadístico sobre algunos de los aspectos que se valoran en el proceso de la cata de un vino y varias de las características que se analizan en los laboratorios enológicos.

La base de datos escogida para este estudio se ha obtenido de la plataforma Kaggle, la cual contiene información detallada sobre la puntuación de diferentes características de vinos blancos y tintos. Está formada por 40 variables y 6497 observaciones, proporcionando así una amplia muestra para el análisis de los datos. Con dicha base de datos se va a investigar cómo influyen las puntuaciones de los diferentes componentes del vino en la puntuación final que se les asignan en las catas.

El análisis descriptivo que se realiza en R (R Core Team 2023) pretende identificar patrones y relaciones significativas entre las diferentes características del vino y su impacto en la puntuación final obtenida en las catas. Esta información es imprescindible para las empresas vinícolas, ya que les permite conocer qué componentes del vino son más determinantes para su calidad y cómo pueden realizar ajustes en el proceso de elaboración para mejorar sus productos.

Las 40 variables mencionadas anteriormente que se van a utilizar para la realización del análisis se dividen en dos tipos, numéricas y categóricas.

#### **Variables numéricas:**

- *fixed.acidity*: la variable acidez fija se refiere a la cantidad de ácidos que no se evaporan con facilidad, presentes en el vino, según los laboratorios enológicos.
- *volatile.acidity*: representa la cantidad de ácidos volátiles presentes en el vino, según los laboratorios enológicos.



- *citric.acid*: indica la cantidad de ácido cítrico que existente en el vino, según los laboratorios enológicos.
- *residual.sugar*: la variable de azúcar residual indica, según los laboratorios enológicos, la cantidad de azúcar que queda en el producto después de que haya terminado el proceso de fermentación.
- *chlorides*: hace referencia a la cantidad de cloruro (sales que contienen cloro y otros elementos) presente en el vino, según los laboratorios enológicos.
- *free.sulfur.dioxide*: representa la cantidad, según los laboratorios enológicos, de dióxido de azufre libre que presenta el vino.
- *total.sulfur.dioxide*: se refiere a la suma de las formas libre y combinada de dióxido de azufre presentes en el vino, según los laboratorios enológicos.
- *density*: la variable densidad se refiere a la consistencia y espesor que tiene el vino, medida en los laboratorios enológicos.
- *pH*: la variable de pH indica si el vino es más ácido o alcalino dependiendo del número que los laboratorios enológicos detecten dentro de la escala establecida.
- *sulphates*: los sulfatos son compuestos químicos derivados del dióxido de azufre (SO<sub>2</sub>) que se pueden encontrar en el vino.
- *alcohol*: representa a la cantidad de alcohol que contiene el vino según los laboratorios enológicos.
- *quality*: la variable calidad indica, a nivel general, la percepción y opinión que tienen los catadores sobre el vino. Está representada por una escala de enteros del 0 al 10.
- *num.reviews*: El número de reseñas se refiere a la cantidad de catas que se han realizado sobre un mismo vino.
- *Flower*: hace referencia a la percepción a florada que el catador aprecia.
- *Plante*: se refiere al porte o apariencia del vino percibida en el proceso de la cata. En este caso y de manera excepcional, la puntuación se ha realizado a la inversa, siendo mejor valorado un número menor.
- *Balance*: es el equilibrio armonioso de los diferentes componentes del vino que se percibe en la cata.
- *Smooth*: es la variable que representa la sensación de suavidad en la boca del catador al beber el vino.
- *Intensity*: hace referencia a la fuerza o potencia de atributos como el olor o el sabor en el vino, de manera generalizada.
- *Odor.Intensity.before.shaking*: es la fuerza o intensidad del olor antes de mover el vino, percibida por el catador.

- *Aroma.quality.before.shaking*: se refiere a la percepción de la calidad del aroma del vino antes de moverlo en la copa.
- *Fruity.before.shaking*: describe la presencia de aromas frutales percibidas por el catador en el vino antes de mover la copa.
- *Flower.before.shaking*: representa la presencia de aromas florales en el vino antes de moverlo en la copa.
- *Spice.before.shaking*: es la presencia de los aromas especiados en el vino antes de agitarlo en la copa, percibidos durante el proceso de la cata.
- *Visual.intensity*: la variable intensidad visual hace referencia al grado de color y claridad del vino que observa el catador en la copa.
- *Nuance*: indica el grado de evolución del vino, es decir, su vejez, si es joven o viejo.
- *Surface.feeling*: hace referencia a la percepción táctil en la boca al beber el vino, relacionada con la textura y la suavidad.
- *Odor.Intensity*: indica la intensidad del olor ,percibido durante el proceso de la cata del vino, después de agitarlo en la copa.
- *Quality.of.odour*: la variable se refiere a la percepción del catador sobre la calidad del olor del vino después de agitarlo en la copa.
- *Fruity*: la variable afrutado describe la presencia de aromas y sabores frutales en el vino.
- *Spice*: se refiere a la presencia de aromas y sabores especiados en el vino.
- *Phenolic*: son los compuestos fenólicos presentes en el vino percibidos en el proceso de cata.
- *Aroma.intensity*: la intensidad del aroma es la fuerza o potencia del aroma percibido por el catador en el vino.
- *Aroma.persistency*: la variable persistencia del aroma es la duración o permanencia del mismo en la boca del catador después de tragar o escupir el vino.
- *Aroma.quality*: esta variable se refiere a la percepción del catador sobre la calidad de los aromas presentes en el vino.
- *Attack.intensity*: se refiere a la intensidad de los sabores y sensaciones del vino al primer contacto en la boca.
- *Acidity*: acidez es la presencia de la sensación de ácido percibido por el catador en el vino.
- *Astringency*: referente a la sensación de sequedad y rugosidad en la boca al consumir el vino.
- *Alcohol*: identifica la sensación percibida del alcohol al consumir el vino.
- *Bitterness*: Es la percepción del sabor amargo en el vino.

- *Harmony*: hace referencia a la percepción de la combinación equilibrada y agradable de diferentes componentes del vino.

### Variables categóricas:

- *class*: indica la variedad de vino al que se refieren las puntuaciones obtenidas en la cata. Este puede tomar los valores de tinto o blanco.

#### 2.1.1 Tecnología

R (R Core Team 2023) ha sido el software escogido para realizar el análisis y estudio de esta base de datos dado que es considerado como la *lingua franca* de la estadística desde hace varios años. Para poder desarrollar el proyecto de una forma más sencilla e intuitiva, se ha utilizado el entorno de desarrollo integrado RStudio (RStudio Team 2023). Además de los paquetes base que incluye el lenguaje, existen muchos otros que facilitan el código y el estudio. Algunos de estos paquetes adicionales que se han utilizado para la realización de este trabajo son los siguientes:

- *clickR*: es un paquete diseñado para corregir de forma semiautomática los errores más comunes presentes en las bases de datos. Esto incluye tanto errores estructurales, como de asignación correcta de clases a las variables, errores tipográficos, distintos formatos de fechas, etc. Adicionalmente, el paquete guarda un registro de todos los cambios realizados, por lo que permite una total reproducibilidad y seguimiento de las modificaciones así como la restauración de los cambios no deseados de forma sencilla y rápida. Facilita el proceso inicial de depuración y puesta a punto de los datos para que el usuario pueda realizarlo de una forma sencilla y fiable (Marin 2023).
- *corr*: el paquete *corr* facilita al usuario la realización y manipulación de matrices de correlación para que así se identifiquen de forma más sencilla y rápida las relaciones significativas entre las variables. También permite elegir entre diferentes métodos de visualización de la matriz, como gráficos de red, mapas de calor, y otras herramientas para personalizar el análisis, que permiten al usuario ver claramente los resultados (Kuhn, S. Jackson y Cimentada 2022).
- *ggplot2*: este paquete es ideal para la realización de gráficos estadísticos. Aunque el paquete base de R (R Core Team 2023) para gráficos es muy bueno y permite crear gráficos de calidad profesional, su sintaxis no sigue unas normas claramente definidas y puede llegar a ser difícil de utilizar para ciertos tipos de gráficos complejos. *ggplot2* sigue los principios de La Gramática de los Gráficos (Wilkinson 2012), por lo que toda su sintaxis es intuitiva y se rige por unas normas claramente definidas y estructuradas, lo que permite la realización de gráficos muy complejos y completamente personalizados con muy poco esfuerzo por parte del usuario (Wickham 2016).
- *ggeffects*: es un paquete diseñado para generar de forma automática efectos marginales y predicciones de modelos estadísticos como por ejemplo regresiones lineales, logísticas y de Poisson. También incluye la creación de gráficos de efectos marginales, permitiendo a los usuarios la personalización y fácil ajuste de estos para mejorar la visualización y la interpretación de los resultados, sin necesidad de centrarse en la complejidad del código (Lüdtke 2018).

- *ggraph*: es una extensión del paquete *ggplot2* que permite al usuario versatilidad y facilidad con la creación de diagramas de alta calidad que representan estructuras complejas. Dispone de diversos métodos de disposición y algoritmos que permiten adaptarse a diferentes tipos de gráficos, ofreciendo un alto nivel de personalización (Pedersen 2024).
- *igraph*: el paquete permite analizar la estructura y propiedades de gráficos complejos gracias a que este proporciona un gran conjunto de funciones diferentes. Es capaz de crear el gráfico partiendo de diversos tipos de fuentes de datos y explorar sus propiedades con facilidad para el usuario, independientemente de si tiene este un gran tamaño (Csárdi et al. 2024).
- *ordinal*: es un paquete diseñado para el análisis de datos en la que su variable respuesta es ordinal, y por tanto se le aplica modelos de regresión ordinal. Incluye funciones para ajustar los modelos de regresión logística ordinal, como el modelo de probabilidad proporcional (modelo logit ordinal), que es perfecto para datos donde las categorías tienen un orden natural pero no una distancia cuantificable entre ellas (Christensen 2023).
- *repmud*: este paquete permite estandarizar las tablas resumen de los modelos independientemente del tipo de modelo y el paquete de R (R Core Team 2023) utilizado para ajustarlo. Por defecto, las tablas resumen muestran las estimaciones de los coeficientes, el error estándar, el intervalo de confianza al 95 % y el p-valor, así como una medida de ajuste relativo del modelo (que suele ser el  $R^2$ , en el caso de modelos lineales, o el criterio de Akaike (AIC) en el caso de otros modelos). En el caso de modelos con función de enlace *logit* o *log*, la tabla resumen muestra también el exponente de los coeficientes, además estima el rendimiento de los diferentes modelos a los que se aplica el paquete, mediante la validación cruzada y bootstrap (Marin 2021).

## 2.2 Modelo de regresión lineal

El modelo de regresión lineal clásico se define como la función de una recta, en la que sus componentes son los siguientes:

$$y = \beta_0 + \sum_{j=1}^p \beta_j X_j + U \quad (2.1)$$

- La variable dependiente ( $y$ ) o variable respuesta, es una función lineal de un conjunto de variables independientes, además se caracteriza por ser continua y distribuida con varianza constante. El valor de esta va a depender del valor que tenga el componente sistemático, es decir los valores que adquieren las diferentes  $X$ 's junto con el error que conlleva.
- $\beta_0$  es el término independiente de la función, es decir, el valor que obtiene  $y$  cuando todas las variables predictoras valen 0.
- $\beta_i$  son los diferentes coeficientes que tienen las variables explicativas, y se interpretan como la pendiente que tiene la recta.

- Las variables  $X_i$  hacen referencia a las diferentes variables independientes, o variables predictoras o explicativas, que forman el modelo, las cuales pueden ser continuas, categóricas o una combinación de ambas.
- $U$  es el término de error que sigue una distribución normal, con media 0 y desviación típica  $\sigma$ .

El análisis de regresión busca identificar un conjunto de variables independientes ( $X_i$ ) que combinadas linealmente, expliquen de manera significativa la variabilidad de una variable dependiente o respuesta  $y$ . Esta capacidad de explicación se evalúa a través de la bondad de ajuste del modelo. Los parámetros de regresión,  $\beta_i$ , indican la influencia individual de cada variable independiente sobre  $y$ , por lo que, si el valor de  $\beta_i$  es elevado, significa que si se produce un cambio en  $X_i$ , el valor esperado de  $y$  variará considerablemente. Si esta variación es grande en relación con la variación del error, significa que es un modelo útil para predecir valores futuros de  $y$  para valores dados de las variables independientes, además de para comprender la importancia relativa de las diferentes variables independientes (George Henry Dunteman 2005).

Este modelo ha sido utilizado desde su creación, pero a en los últimos años, los científicos han reconocido que el modelo de regresión lineal tiene una serie de limitaciones. Una de ellas es que, como se ha mencionado anteriormente, la variable dependiente ha de ser continua o cuasicontinua, pero en ocasiones esta ha pasado a ser categórica y por tanto incompatible con este modelo. Como solución se creó un nuevo modelo idóneo para estos casos, el modelo lineal generalizado (George Henry Dunteman 2005).

Existen diferentes modelos de regresión pertenecientes a los tipos de modelos lineales generalizados, caracterizados porque su variable dependiente es no normal (George Henry Dunteman 2005). Estos modelos implican predecir el valor de una variable dependiente como una función lineal de un conjunto de variables independientes, con el objetivo de modelar la probabilidad de que ocurra el evento de interés.

Los dos componentes de un modelo lineal generalizado son:

- La función de enlace, que es una transformación de la media de la variable dependiente de modo que esta variable transformada es una función lineal de los parámetros de regresión.
- Una distribución de error.

## 2.3 Regresión ordinal

La regresión logística ordinal es uno de los tipos de los modelos lineales generalizados explicados anteriormente. Esta regresión es un método que se utiliza para modelar la relación entre una variable respuesta de tipo ordinal y una o más variables independientes. La variable ordinal o respuesta, es de tipo categórica y por tanto, para ella existe un orden claro de los niveles de categoría, mientras que las variables explicativas que forman el modelo pueden ser tanto continuas como categóricas (Parry 2010).

Para que esto suceda, el modelo aplica una función de enlace canónica para transformar los resultados obtenidos de los parámetros de regresión, en una función lineal (George Henry Dunteman 2005). En el caso de la regresión ordinal, la función canónica que se aplica es *logit*, y la regresión se expresa de la siguiente forma según lo explicado anteriormente:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2.2)$$

- Donde el resultado obtenido es  $\log\left(\frac{p}{1-p}\right)$ , que a su vez es la definición de *logit*( $p$ ), y  $\beta_p X_p$  son los coeficientes de regresión.

Utilizando esta función de enlace canónica, la interpretación de la variable respuesta cambia con respecto a la realizada en los modelos de regresión lineal, puesto que en esta regresión lo que se modela es la probabilidad de que suceda el evento de interés ( $y$ ), y por tanto, siendo este cambio el punto de unión entre la variable respuesta  $y$  y el valor obtenido en  $\log\left(\frac{p}{1-p}\right)$  (George Henry Dunteman 2005).

Para poder estimar la calidad de los modelos de regresión ordinal creados, no es posible la utilización de las técnicas que se aplican en la regresión lineal.

El AIC es aplicable a una gran variedad de modelos, como son la regresión lineal, regresión ordinal, series temporales, etc. Por este motivo, el AIC ha sido el criterio escogido para identificar el mejor modelo en este Trabajo Fin de Grado.

El AIC (Akaike Information Criterion o Criterio de Información de Akaike) es una medida de la calidad relativa de un modelo estadístico, para así poder identificar la idoneidad del mismo en función de sus componentes. Se basa en la medición de la precisión usada en los dos criterios que se han de tener en cuenta a la hora de evaluar un modelo para escogerlo como definitivo, la precisión o bondad del ajuste y la escasez de variables o simplicidad del modelo.

Como bien se ha comentado anteriormente, se evalúan ambos criterios de manera objetiva y se busca un equilibrio entre ambos. Es decir, la introducción de nuevas variables en el modelo únicamente será valorada positivamente por el AIC cuando esta produzca una mejora en la bondad de ajuste lo suficientemente relevante (Dunn, Smyth et al. 2018). Cabe destacar que el AIC con menor valor explica que el modelo es más idóneo y ajustado, debido a la fórmula que lo define:

$$AIC = 2k - 2\ln(L) \quad (2.3)$$

Donde:

- $k$  es el número de parámetros existentes en el modelo, por lo que un valor elevado penaliza el modelo por existir sobreajuste.
- $\ln(L)$  es el logaritmo natural de la verosimilitud (log-likelihood) del modelo, es decir, cómo de idóneo es que la variable pertenezca al modelo.



# Resultados

### 3.1 Análisis Descriptivo

A continuación, se muestran los resultados obtenidos del análisis descriptivo que se ha llevado a cabo para conocer con más profundidad las variables de la base de datos y los valores que tienen. Para ello se han utilizado las diferentes librerías del software R (R Core Team 2023) que se han mencionado en el apartado anterior 2.1 Obtención de la información.

Para poder proporcionar un resumen general y comprensivo de las características y la distribución de las distintas variables que forman la base de datos y observar si existe algún valor que se desvía significativamente (outlier), se ha realizado el cálculo de diferentes estadísticos como la media, mediana y primer y tercer cuartiles. Esto se muestra en la tabla 3.1. Cabe destacar que, para la variable *quality* no se pueden obtener estos estadísticos por ser una variable categórica, por lo que en la tabla 3.1 se muestra la frecuencia con la que aparecen los valores de dicha variable.

- La media proporciona una idea del valor central de la variable sobre la que se calcula, y es el promedio de los valores de dicha variable.
- La mediana es el valor central que se obtiene cuando los datos están ordenados. Es menos sensible a los valores atípicos que la media.
- Los cuartiles dividen los datos en cuatro partes para conocer la dispersión existente en los datos y con ello si existe una concentración de valores en algún cuartil. El primer cuartil representa el 25 % de los datos que están por debajo de este valor. El segundo cuartil coincide con el valor de la mediana ya que este representa la mitad de los datos, mientras que el tercero, representa el 75 % de los datos ubicados por debajo del valor.

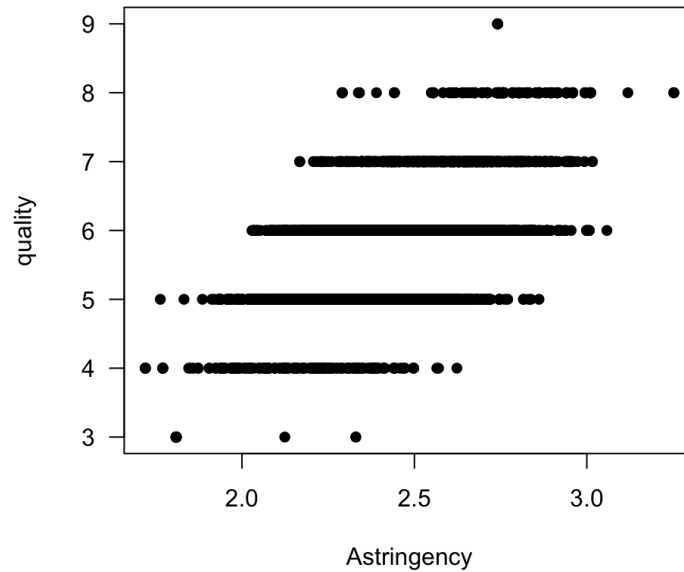
Dado que la variable objeto de estudio, como bien se ha comentado anteriormente es “quality”, se ha querido observar la correlación existente de la variable respuesta con algunas de las variables independientes.



Variable	Media (SD)	Mediana (1er, 3er cuartil)	n = 6497
fixed.acidity	7.22 (1.3)	7 (6.4, 7.7)	
volatile.acidity	0.34 (0.16)	0.29 (0.23, 0.4)	
citric.acid	0.32 (0.15)	0.31 (0.25, 0.39)	
residual.sugar	5.44 (4.76)	3 (1.8, 8.1)	
chlorides	0.06 (0.04)	0.05 (0.04, 0.06)	
free.sulfur.dioxide	30.53 (17.75)	29 (17, 41)	
total.sulfur.dioxide	115.74 (56.52)	118 (77, 156)	
density	0.99 (0)	0.99 (0.99, 1)	
pH	3.22 (0.16)	3.21 (3.11, 3.32)	
sulphates	0.53 (0.15)	0.51 (0.43, 0.6)	
alcohol	10.49 (1.19)	10.3 (9.5, 11.3)	
num.reviews	453.45 (668.43)	404 (389, 415)	
Flower	2.17 (0.25)	2.16 (2, 2.33)	
Plante	1.96 (0.2)	1.96 (1.82, 2.1)	
Balance	3.2 (0.35)	3.22 (2.93, 3.4)	
Smooth	2.67 (0.45)	2.67 (2.34, 2.94)	
Intensity	3.17 (0.4)	3.16 (2.89, 3.44)	
Odor.Intensity.before.shaking	3.12 (0.29)	3.12 (2.92, 3.31)	
Aroma.quality.before.shaking	3.05 (0.2)	3.05 (2.92, 3.18)	
Fruity.before.shaking	2.72 (0.2)	2.72 (2.58, 2.86)	
Flower.before.shaking	2.06 (0.14)	2.06 (1.96, 2.15)	
Spice.before.shaking	2 (0.25)	2 (1.82, 2.17)	
Visual.intensity	3.97 (0.53)	3.96 (3.61, 4.32)	
Nuance	3.74 (0.51)	3.73 (3.4, 4.08)	
Surface.feeling	3.19 (0.27)	3.18 (3, 3.37)	
Odor.Intensity	3.4 (0.2)	3.4 (3.26, 3.53)	
Quality.of.odour	3.24 (0.22)	3.24 (3.08, 3.4)	
Fruity	2.85 (0.22)	2.85 (2.7, 3)	
Spice	2.18 (0.21)	2.18 (2.03, 2.32)	
Phenolic	1.54 (0.15)	1.55 (1.44, 1.65)	
Aroma.intensity	3.21 (0.29)	3.2 (3.03, 3.4)	
Aroma.persistency	2.99 (0.27)	2.99 (2.81, 3.16)	
Aroma.quality	3.08 (0.35)	3.07 (2.83, 3.31)	
Attack.intensity	3.17 (0.34)	3.16 (2.94, 3.38)	
Acidity	2.37 (0.24)	2.37 (2.21, 2.53)	
Astringency	2.45 (0.21)	2.44 (2.31, 2.59)	
Alcohol	2.75 (0.18)	2.74 (2.62, 2.87)	
Bitterness	2.07 (0.19)	2.07 (1.94, 2.2)	
Harmony	3.17 (0.51)	3.18 (2.8, 3.47)	
quality			
3	30 (0.46 %)		
4	216 (3.32 %)		
5	2138 (32.91 %)		
6	2836 (43.65 %)		
7	1079 (16.61 %)		
8	193 (2.97 %)		
9	5 (0.08 %)		

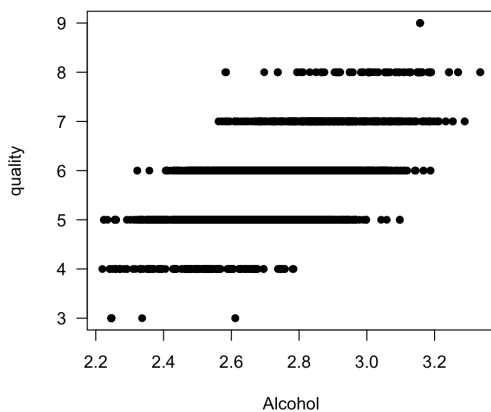
**Tabla 3.1:** Análisis descriptivo de la base de datos.

La figura 3.1 muestra la relación existente entre la variable respuesta calidad y la variable independiente astringencia. Se observa que existe una tendencia ascendente en la que a mejor puntuación general se le ha dado al vino, mayor astringencia se ha percibido en la cata, hasta un cierto punto, dado que los vinos con mejor puntuación extrañamente no sobrepasan el valor cercano a 2.9. En cambio, los pocos vinos que se han puntuado con una calidad general de 9, tienen una astringencia cercana al 2.7.

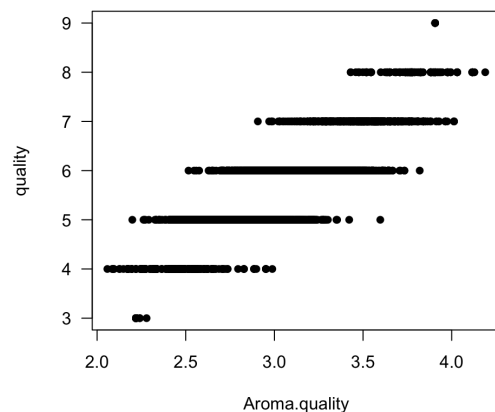


**Figura 3.1:** Correlación entre las variables calidad y astringencia.

También se observa en la figura 3.2 y figura 3.3 que, en ambos casos, la calidad aumenta a medida que lo hace el alcohol percibido al tomar el vino, al igual que aumenta si se valora positivamente la calidad del aroma.

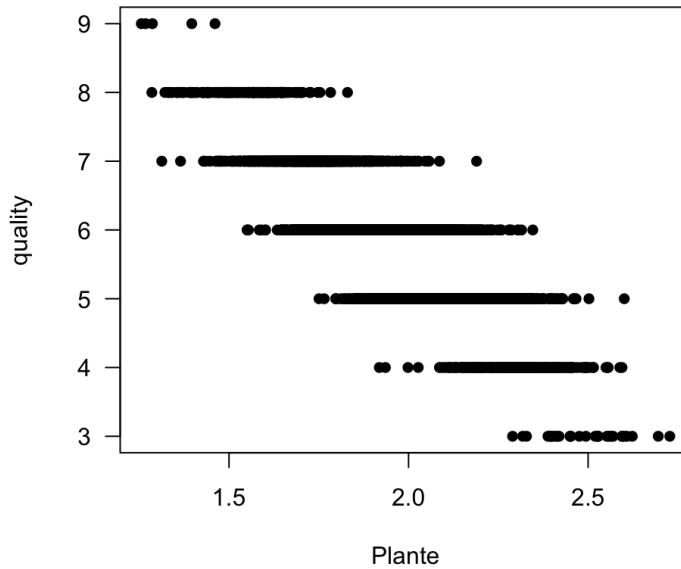


**Figura 3.2:** Correlación entre la calidad y el alcohol percibido.



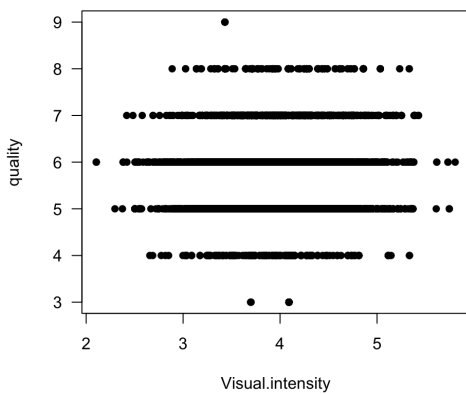
**Figura 3.3:** Correlación entre la calidad y la variable calidad del aroma percibida.

Claramente se observa en la figura 3.4 una relación existente entre la variable planta y la calidad del vino, y se aprecia que es descendente. Esto es causado por los motivos ya explicados en la descripción de variables, lo que indica que a una mejor valoración, mejor porte tiene el vino.

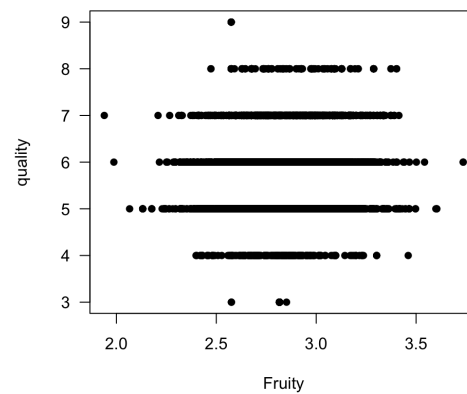


**Figura 3.4:** Correlación entre la calidad y la planta (porte) del vino.

Cabe destacar que, a pesar de que a priori podría asumirse que existe una relación entre los datos de la calidad del vino y la intensidad visual del mismo, no se ha encontrado ninguna, al igual que entre la cantidad de sabor percibida en el vino y su calidad global (figura 3.5, figura 3.6).



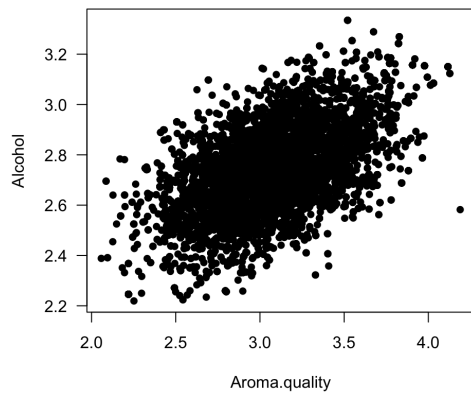
**Figura 3.5:** Correlación entre la calidad y la intensidad visual percibida del vino.



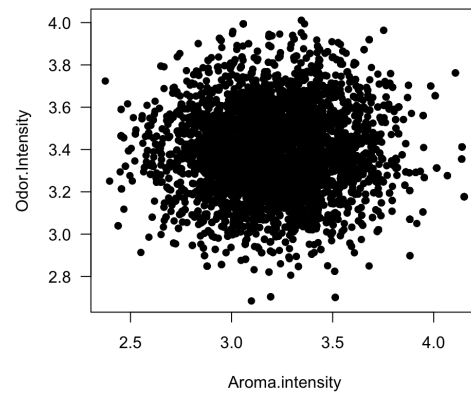
**Figura 3.6:** Correlación entre la variable calidad y sabrosidad del vino percibidos.

Por otro lado, es destacable comentar las relaciones existentes entre las variables independientes siguientes. Pues existe una relación positiva entre las variables alcohol percibido e intensidad del aroma también percibido, tal y como se ha explicado en el apartado 1.3.2 Aromas y alcohol (figura 3.7), al igual que se comenta en ese apartado la correlación negativa que se muestra en la figura 3.9 de la densidad del vino y el alcohol que se percibe.

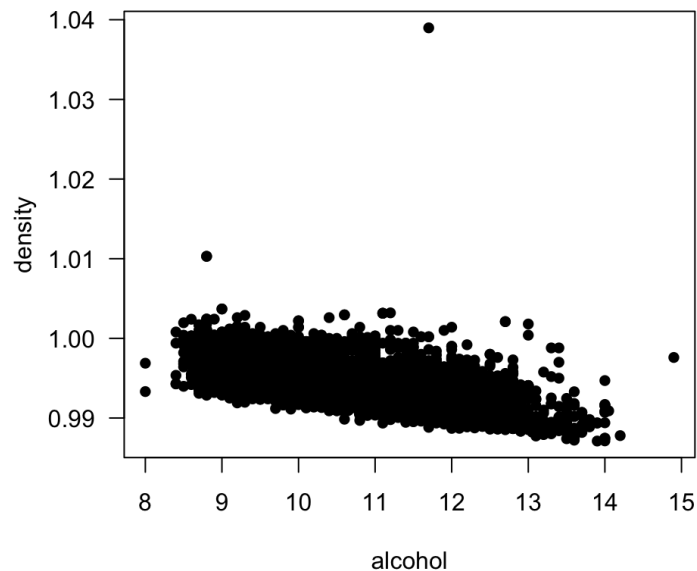
Por lo contrario, no es posible observar una relación entre la intensidad del olor y la intensidad del aroma que el catador percibe a pesar de que a priori era lógico suponer la existencia de dicha relación (figura 3.8).



**Figura 3.7:** Correlación entre el alcohol y la intensidad del aroma percibidos.



**Figura 3.8:** Correlación entre la intensidad del olor percibido y la intensidad aroma percibida.

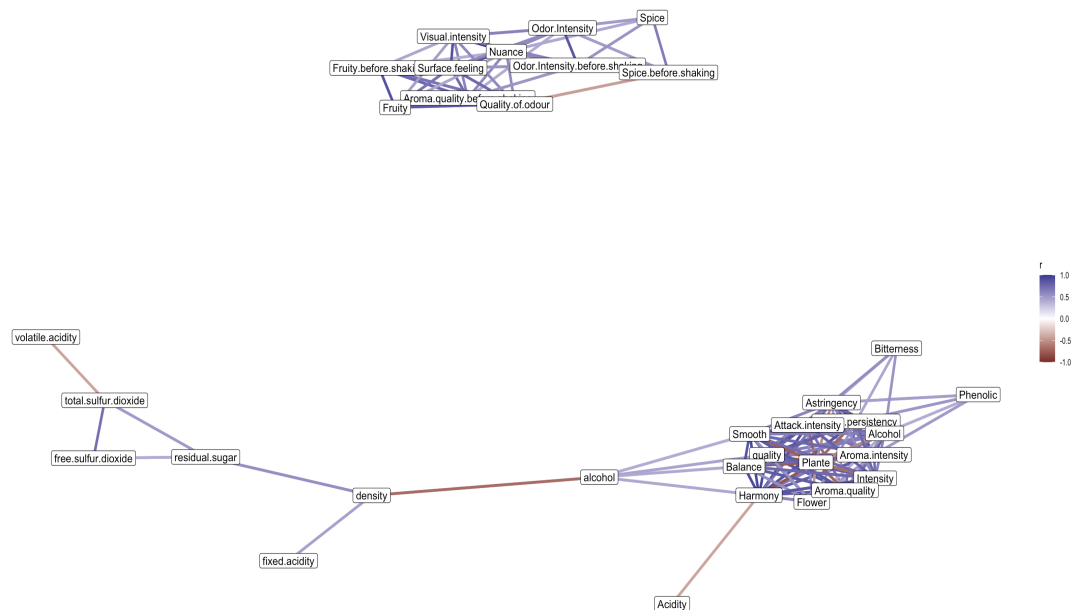


**Figura 3.9:** Correlación entre la variable densidad del vino y alcohol que contiene el mismo.

Una vez se ha destacado la existencia, o no, de estas relaciones, mediante el algoritmo de diseño de Fruchterman-Reingold (Fruchterman y Reingold 1991) se ha elaborado el gráfico correspondiente a la matriz de correlación, estudiando las diferentes conexiones existentes entre todas las variables de la base de datos que su valor de correlación es igual o superior al indicado, excluyendo así las relaciones que estuviesen por debajo de dicha cifra.

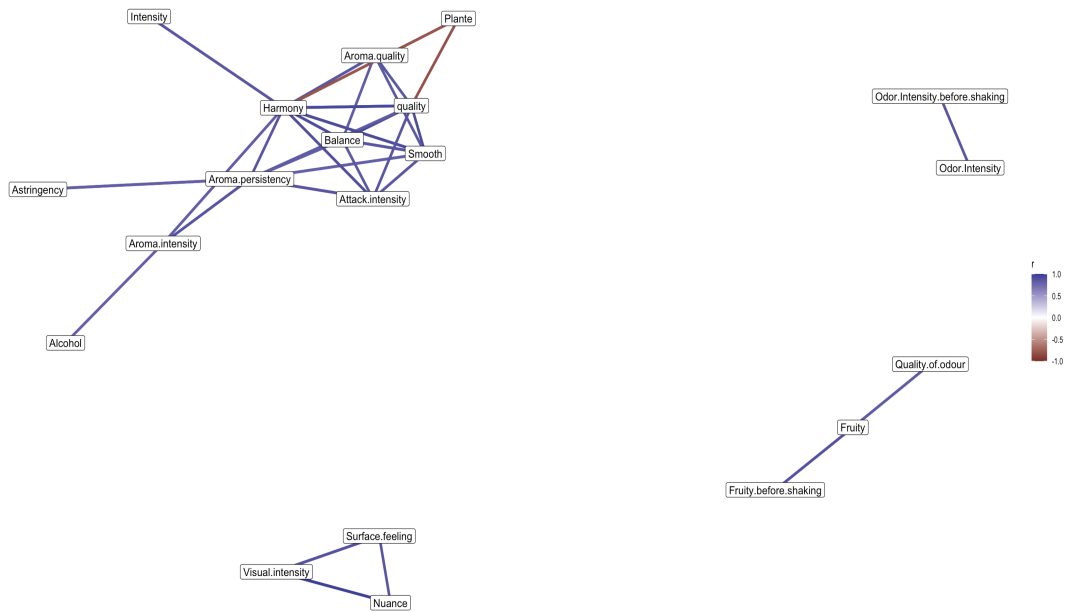
Gracias a este tipo de gráficos es posible observar de forma sencilla las diferentes relaciones que tiene cada una de las variables frente al resto de ellas y según el color del arco que las une se indica si la relación es positiva (color morado) o negativa (color magenta) y según si este color es más o menos intenso, se indica si el valor de correlación es mayor o menor respectivamente. La figura 3.10 muestra estas conexiones para una correlación entre variables de mínimo 0.4.

En ella se puede observar que existen tres grupos de variables relacionadas entre sí, de los cuales uno de ellos, ubicado en la parte central superior de la imagen, no interacciona con los otros dos grupos. También es posible ver las relaciones mencionadas anteriormente en el apartado 1.3 Componentes, como la del alcohol con la densidad, pero también se observa que otras relaciones ya mencionadas, no se muestran por tener valores inferiores a 0.4, como por ejemplo la relación entre la calidad y la intensidad visual.



**Figura 3.10:** Matriz de correlación para correlaciones superiores a 0.4, según el algoritmo de diseño Fruchterman-Reingold.

Para poder observar con mayor claridad las relaciones entre variables más fuertes, se ha establecido un valor de correlación de 0.8. Esto se muestra en la figura 3.11, donde se muestra que las relaciones han disminuido respecto a la figura 3.10 y de los tres grupos que se observaban, tan solo queda uno mayormente visible, y los otros dos se han ido deshaciendo hasta quedar solo un par de variables relacionadas.



**Figura 3.11:** Matriz de correlación para correlaciones superiores a 0.8, según el algoritmo de diseño Fruchterman-Reingold.

## 3.2 Desarrollo del modelo

Tras realizar el análisis descriptivo, se conoce con mayor precisión las variables de las que se dispone para la elaboración del modelo. Con este conocimiento adquirido, se ajustan diversos modelos compuestos por una combinación diferente de variables y mediante el uso de la herramienta AIC, explicada con detalle en el apartado 2.3 Metodología, se selecciona el modelo que tenga el AIC menor y que por tanto sus variables aporten más información.

El primer modelo que se ha realizado tiene en cuenta algunas de las variables que se habían observado de forma más individualizada en el apartado anterior (3.1), como Plante, y se compone de la siguiente forma expresada en lenguaje R:

```
model1 <- clm(quality ~ Aroma.persistency + Balance + density + Plante, data = datos_vinos)
```

Si se ejecuta el comando *summary* para el *model1*, se puede observar en la tabla 3.2 que el AIC correspondiente al modelo, tiene un valor de 1317.61. Este valor no tiene mucho significado por sí mismo, ya que lo que importa es la diferencia entre los valores de AIC de varios modelos, por lo que se va a crear un nuevo modelo, partiendo del anterior, para observar si se mejora la medida.

El *model2* mantiene las variables que componen al modelo anterior y para modificarlo, se le han añadido las nuevas variables *Acidity* y *Phenolic* y así poder observar su efecto en el valor del AIC. Como se observa en la tabla 3.3, este ha disminuido significativamente hasta alcanzar el valor de 1185.31, significando que el *model2* se ajusta mejor a los datos que el *model1*.

	Estimate	Std. Error	Lower 95 %	Upper 95 %	P value
Aroma.quality	4.144	0.424	3.339	4.974	<0.001
Aroma.persistency	7.038	0.516	422.516	3195.298	<0.001
Balance	29.586	1.044	983255758268.349	59045840060038.1	<0.001
density	-125.816	23.092	0	0	<0.001
Plante	-13.79	0.753	0	0	<0.001
Phenolic	1.19	0.514	0.194	2.231	0.022
AIC	1317.611				

**Tabla 3.2:** Resultados obtenidos tras ejecutar el comando `summary(model1)`.

	Estimate	Std. Error	Lower 95 %	Upper 95 %	P value
Aroma.persistency	10.917	0.707	14231.582	227861.473	<0.001
Balance	29.056	1.104	517909887378.252	39473344750034.3	<0.001
Acidity	-4.009	0.4	0.008	0.039	<0.001
density	-134.884	24.483	0	0	<0.001
Plante	-13.537	0.791	0	0	<0.001
Phenolic	-3.104	0.636	0.013	0.155	<0.001
AIC	1185.31				

**Tabla 3.3:** Resultados obtenidos tras ejecutar el comando `summary(model2)`.

Observando la figura 3.11, en el *model3* también se decide cambiar la variable de *Aroma.persistency* por *Aroma.quality*, dado que ambas están relacionadas directamente con la variable respuesta *quality* y es interesante observar cuál es la variación que tiene este cambio en la explicación de la variable respuesta y por tanto, en el valor del AIC. Este cambio ha mejorado el modelo como muestra la tabla 3.4, en 69.2 puntos según la diferencia existente entre el AIC del *model1* y el del *model2*.

	Estimate	Std. Error	Lower 95 %	Upper 95 %	P value
Aroma.quality	8.316	0.507	1552.782	11349.415	<0.001
Balance	30.008	1.173	1183785222717.28	118235132157656	<0.001
Acidity	-1.73	0.357	0.087	0.355	<0.001
density	-142.653	25.142	0	0	<0.001
Plante	-14.042	0.828	0	0	<0.001
Phenolic	3.797	0.58	14.455	140.538	<0.001
AIC	1116.113				

**Tabla 3.4:** Resultados obtenidos tras ejecutar el comando `summary(model3)`.

Viendo los resultados anteriores y las variables utilizadas, es interesante plantear un modelo que tenga ambas variables referentes al aroma, dado que al usarlas, en ambos casos ha mejorado el modelo. Además, en este nuevo *model4* se ha añadido también otra variable que lo modifica, quedando de la siguiente forma según la sintaxis del lenguaje R:

```
model4 ← clm(quality ~ Aroma.quality + Balance + Acidity + density + Plante + Phenolic + Aroma.persistency + Smooth, data = datos_vinos)
```

Estas modificaciones se reflejan en el valor del AIC de manera positiva, dado que, como se observa en la tabla 3.4, el valor del AIC ha disminuido drásticamente hasta alcanzar el valor de 533.91, que comparado con el del *model1* (1317.61), es mucho mejor.

	Estimate	Std. Error	Lower 95 %	Upper 95 %	P value
Aroma.quality	4.385	0.699	20.885	324.678	<0.001
Balance	29.089	1.697	186924777608.118	147011897345772	<0.001
Acidity	-2.685	0.608	0.02	0.22	<0.001
density	-128.541	39.299	0	0	0.001
Plante	-13.182	1.152	0	0	<0.001
Phenolic	0.911	0.983	0.364	17.241	0.354
Aroma.persistency	4.324	1.072	9.555	642.154	<0.001
Smooth	13.032	0.934	79653.386	3124738.554	<0.001
AIC	533.91				

**Tabla 3.5:** Resultados obtenidos tras ejecutar el comando `summary(model4)`.

En cambio el *model4* no es el elegido como modelo definitivo, ya que, de nuevo se ha observado la figura 3.10 para estudiar si existe alguna variable adicional que a priori pueda aportar información al modelo usando los conocimientos adquiridos en el apartado 1.3 Componentes. Tras realizar diversas pruebas en las que se añadían al modelo anterior diferentes variables como *citric.acidity* o *pH*, se ha obtenido un modelo que mejora al ya conocido *model4*. A continuación, se muestra el nuevo y definitivo *model5* formado por la siguiente estructura según la sintaxis de R:

```
model5 ← clm(quality ~ Aroma.quality + Balance + Acidity + density + Plante + Phenolic
+ Aroma.persistency + Smooth + Attack.intensity + chlorides, data = datos_vinos)
```

	Estimate	Std. Error	Lower 95 %	Upper 95 %	P value
Aroma.quality	6.096	0.838	89.994	2415.212	<0.001
Balance	27.871	1.729	52338519121.098	46697069992375.4	<0.001
Acidity	-3.029	0.645	0.013	0.167	<0.001
density	-92.395	46.835	0	0.922	0.049
Plante	-12.665	1.212	0	0	<0.001
Phenolic	3.042	1.129	2.354	197.955	0.007
Aroma.persistency	0.133	1.37	0.077	16.719	0.923
Smooth	11.561	0.97	17024.927	770160.371	<0.001
Attack.intensity	5.892	1.11	43.215	3365.213	<0.001
chlorides	-7.023	2.909	0	0.286	0.016
AIC	501.192				

**Tabla 3.6:** Resultados obtenidos tras ejecutar el comando `summary(model5)`.

Se observa que el *model5* ha mejorado el valor del AIC que se tenía hasta disminuirlo a la cifra de 501.19, tal y como muestra la tabla 3.6 tras ejecutar el comando `summary(model5)`.

En los resultados del modelo mostrados en la tabla 3.6, además de observar el valor del AIC, es interesante destacar que el coeficiente de la variables *density* tienen un valor de *Estimate* muy elevado, lo que significa que en el modelo hay separación perfecta (Lipsitz et al. 2013). Este problema detectado hace referencia a que el algoritmo utilizado no es capaz de encontrar una combinación de valores de coeficientes que minimice el error entre la separación de las clases, y por tanto, los coeficientes que producen esa separación perfecta tienden a  $+\infty$  o a  $-\infty$ .

Para poder corregir dicho problema se ha utilizado el método propuesto por Lipsitz et al. 2013, que consiste en penalizar el ajuste del modelo para conseguir que los coeficientes no alcancen valores tan elevados como los ya mencionados. Tras la elaboración de dichas correcciones, se han obtenido los resultados que se muestran la tabla 3.7.



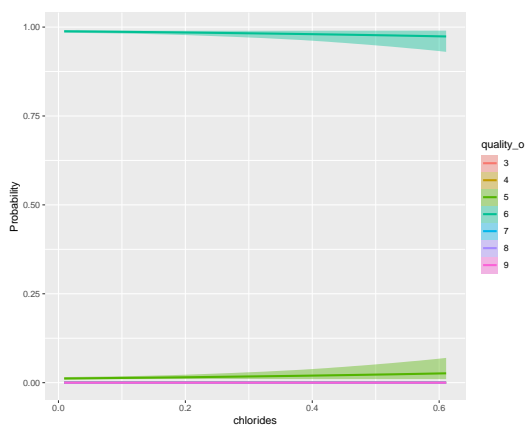
	Estimate	Std. Error	Lower 95 %	Upper 95 %	P value
Intercept[1]	45.508	2.832	40.201	51.142	<0.001
Intercept[2]	57.97	2.947	52.452	63.859	<0.001
Intercept[3]	72.884	3.16	66.936	79.256	<0.001
Intercept[4]	86.889	3.392	80.476	93.817	<0.001
Intercept[5]	100.821	3.659	93.884	108.315	<0.001
Intercept[6]	114.193	3.962	106.685	122.345	<0.001
Aroma.quality	4.144	0.424	3.339	4.974	<0.001
Balance	13.564	0.539	12.523	14.634	<0.001
Acidity	-2.083	0.353	-2.751	-1.427	<0.001
density	-0.153	0.997	-2.158	1.795	0.876
Plante	-6.397	0.55	-7.498	-5.32	<0.001
Phenolic	1.19	0.514	0.194	2.231	0.022
Aroma.persistency	1.606	0.587	0.497	2.735	0.006
Smooth	7.022	0.454	6.142	7.918	<0.001
Attack.intensity	4.216	0.509	3.228	5.198	<0.001
chlorides	-1.348	0.902	-3.107	0.384	0.133
AIC	500.02				

**Tabla 3.7:** Resultados obtenidos tras ejecutar el modelo con la corrección de Lipsitz.

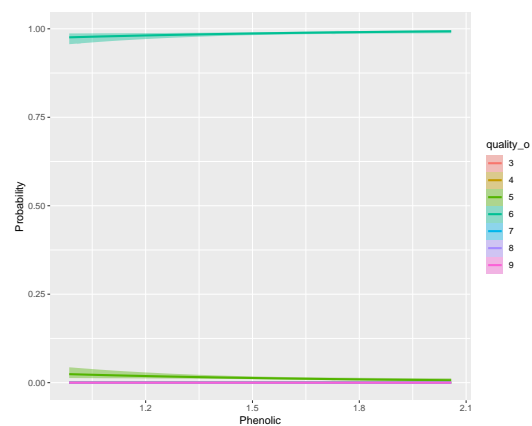
### 3.3 Interpretación

Una vez se ha escogido un modelo como definitivo para ser estudiado, se han de interpretar sus parámetros.

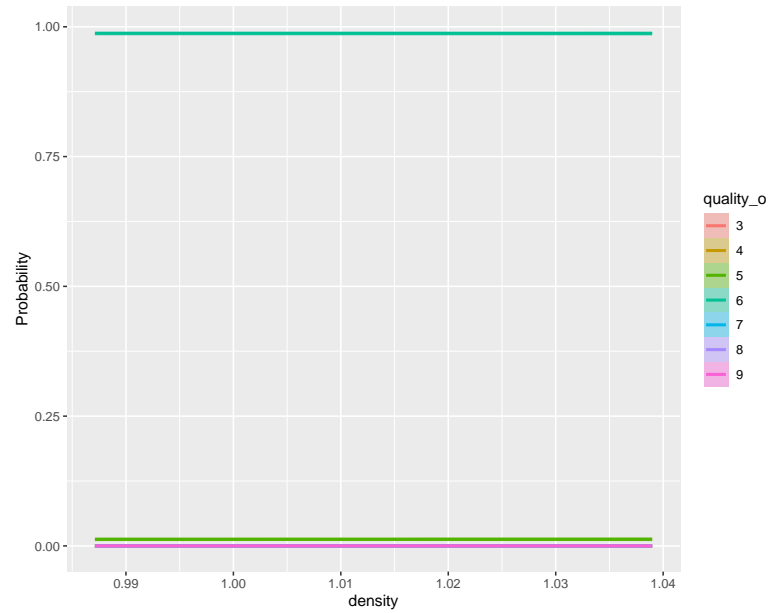
La figura 3.12 , figura 3.13 y figura 3.14 muestran de una forma clara que las variables *chlorides*, *Phenolic* y *density*, especialmente la última de ellas, no tienen efecto significativo sobre la calidad del vino, lo que significa que dichas variables no ayudan a explicar la variable respuesta *quality*. Esto corrobora lo visto en los resultados de la columna *Pvalue* en la tabla 3.7, puesto que dichas variables tienen valores superiores a 0.001.



**Figura 3.12:** Gráfica de efectos para la variable *chlorides*.

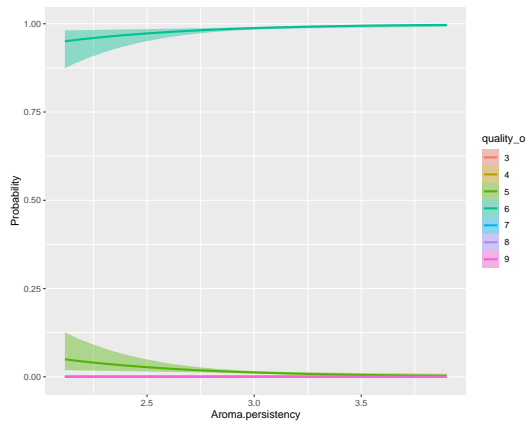


**Figura 3.13:** Gráfica de efectos para la variable *Phenolic*.

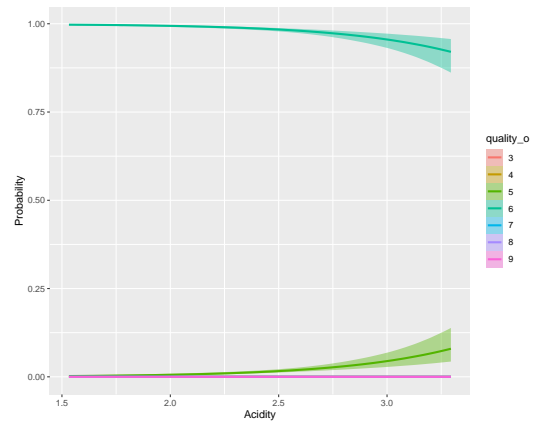


**Figura 3.14:** Gráfica de efectos para la variable *density*.

Por otro lado, la variable *Aroma.persistency* se interpreta en la figura 3.15, como que si se puntúa con valores comprendidos entre aproximadamente el 2 y el 2.75, existe una probabilidad cercana al 95 % de que la calidad del vino se valore con un 6, mientras que el porcentaje restante de probabilidad es de obtener una calidad de 5. Por lo contrario, si se observa la figura 3.16 sucede que si hay una mayor acidez, la probabilidad de obtener una puntuación de 6 en la calidad general del vino disminuye hasta un 87 % aproximadamente (si no se obtendría un 5), mientras que si existe una acidez puntuada con un valor igual i inferior a 2, la probabilidad de obtener un 6 en la calidad del vino es del 100 %.



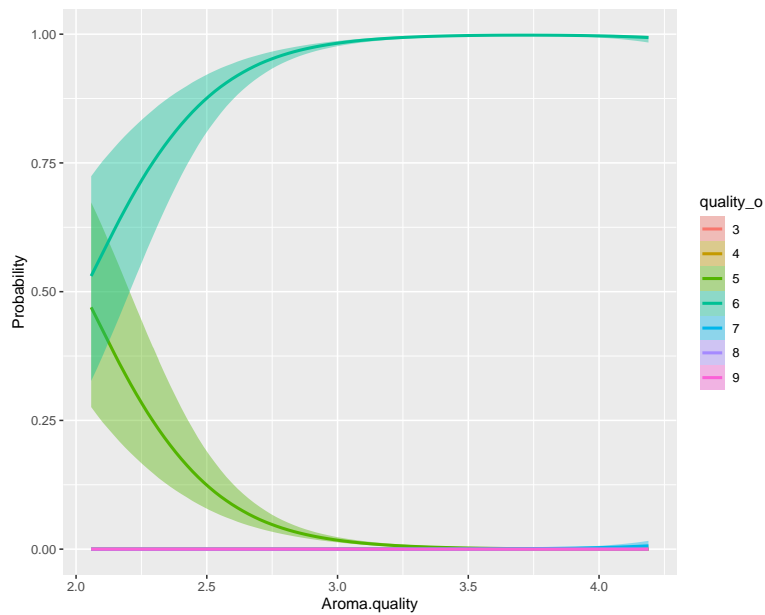
**Figura 3.15:** Gráfica de efectos para la variable *Aroma.persistency*.



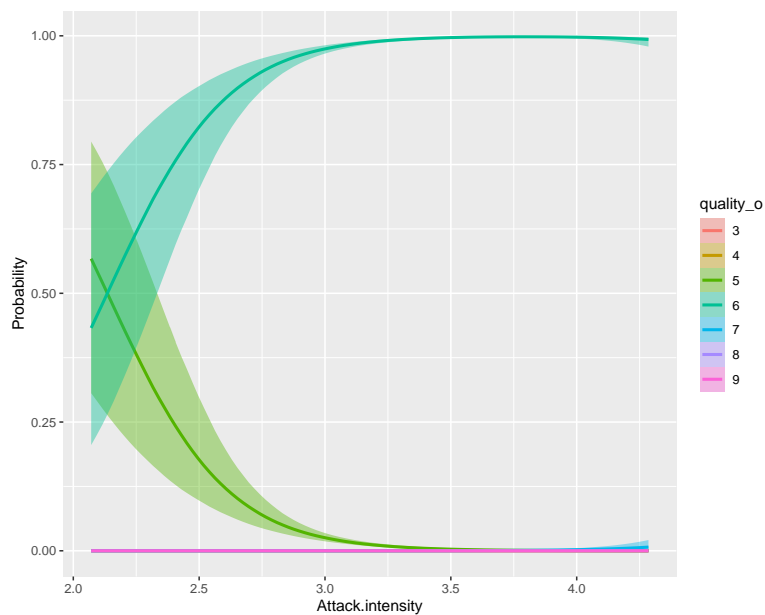
**Figura 3.16:** Gráfica de efectos para la variable *Acidity*.

Existe una probabilidad inferior al 45 %, de que puntuando la calidad del aroma con un valor próximo al 2, se valore la calidad general del vino con un 5. Mientras que, con estos mismos valores para la calidad del aroma, existe una probabilidad de 65 % aproximadamente de que la calificación sea de 6, tal y como se muestra en la figura 3.17, esta probabilidad de obtener un valor de 6 va aumentando a medida que se incrementa la intensidad del sabor percibida, hasta que esta llega a un valor de aproximadamente 4.1, donde aparece una ínfima probabilidad de obtener una puntuación de la calidad de 7.

Lo explicado anteriormente, sucede de igual manera para la variable *Attack.intensity*, (figura 3.18), la cual tiene una probabilidad ligeramente mayor (cerca del 55%) para la obtención del 5 en la variable respuesta y por ende, inferior para la puntuación de 6.



**Figura 3.17:** Gráfica de efectos para la variable *Aroma.quality*.

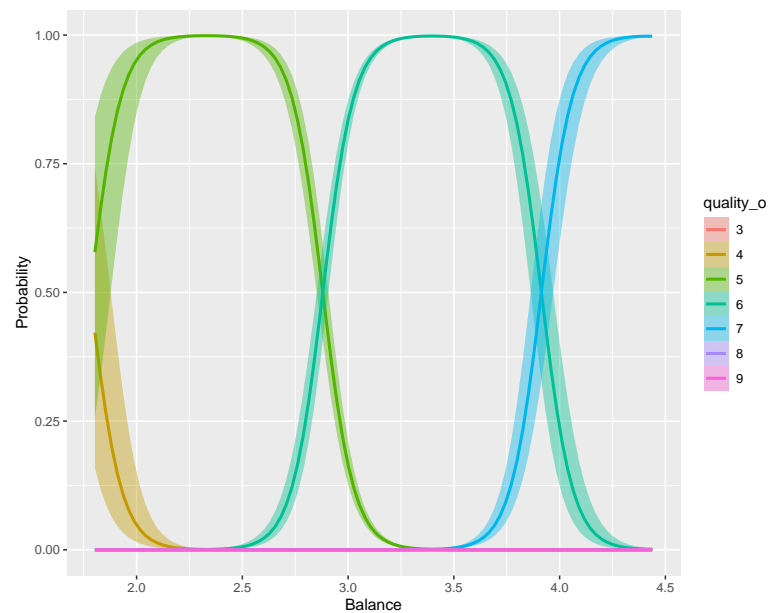


**Figura 3.18:** Gráfica de efectos para la variable *Attack*.

De forma contraria a lo observado en las figuras anteriores, en la figura 3.19 se observa como existe una probabilidad de que se produzca una calificación concreta en la calidad final del vino si se puntua de una determinada forma al balance existente entre los componentes del vino. De forma generalizada, se puede deducir que a mayor puntuación de *Balance*, mejor calidad tiene el vino. Entrando más en detalle, esto se produce porque se observa como la probabilidad, cercana al 40%, que existe de obtener una calidad puntuada como 4, se da cuando en la variable *Balance* no se supera el valor de 2 aproximadamente.

En cambio, la máxima probabilidad de obtener un 5 en la calidad del vino se da si la puntuación del balance es igual o superior a 2.25. La probabilidad de obtener una puntuación de 6 en la calidad del vino, va aumentando exponencialmente desde que se valora la variable *Balance* con un 2.40, hasta que se puntúa con un valor muy cercano a 3, pues a partir de esta cifra, la probabilidad de obtener esa calificación de 6 en la calidad se mantiene constante hasta que el balance toma valores comprendidos entre el 3.5 y el 4.4, que la probabilidad desciende drásticamente hasta llegar a 0.

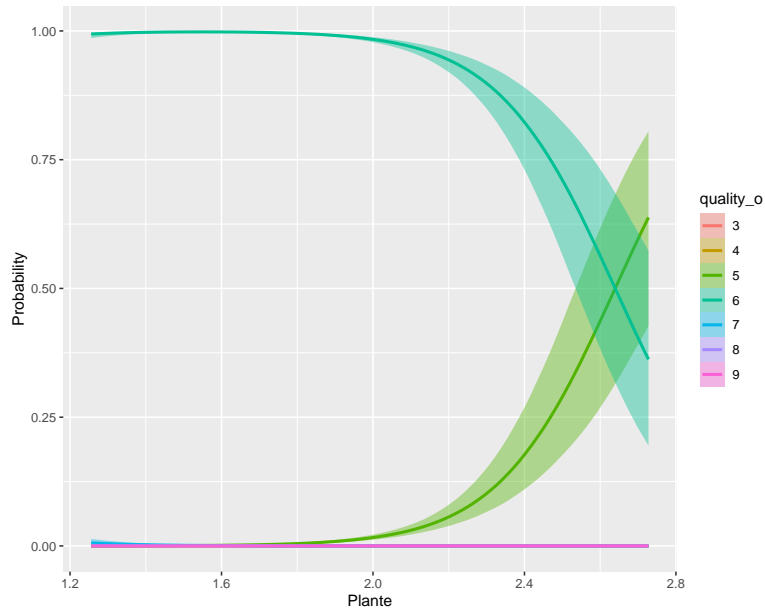
Esto mismo sucede con la probabilidad de obtener un 7 en la variable *quality*, dado que su probabilidad comienza cuando se califica al balance con un valor de 3.5, llegando al 100 % cuando este vale 4.4, pero en cambio, no existe probabilidad de obtener un valor de la calidad superior a 7 si el balance es igual o inferior a 4.5 que es el máximo valor obtenido como puntuación de esta variable.



**Figura 3.19:** Gráfica de efectos para la variable *Balance*.

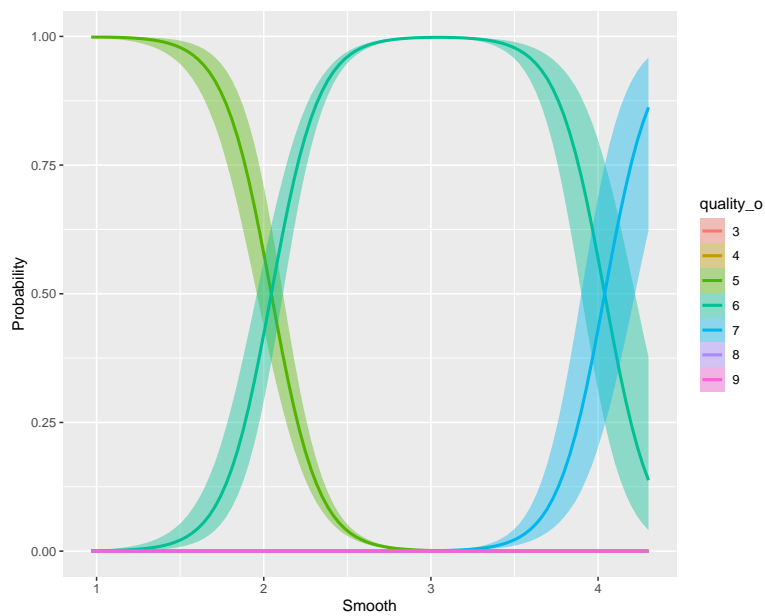
En la figura 3.20 se explica de forma visual que, a mayor valor de la variable *Plante*, menor puntuación de calidad se obtiene, es decir, cuando se valora que el vino tiene menos porte o presencia, la calidad general de este es mayor. Esto es debido a que la probabilidad de obtener una puntuación más elevada, disminuye si la calificación de la planta es mayor y viceversa. Por ejemplo, si se valora el porte con cifras inferiores a 2, la probabilidad de obtener una calificación mayor a 6 disminuye, pero en cambio si se puntúa la planta de forma superior al mismo valor mencionado, la probabilidad de que se puntúe con una calificación de 5, aumenta. En este caso, la interpretación puede llevar a confusión en ciertas ocasiones porque los valores con los que se ha valorado una mejor planta son los menores.

Como se observa en la figura 3.21, no existe probabilidad de que se obtengan calificaciones iguales o superiores al 8 ni iguales o inferiores al 5 para la calidad de un vino de forma generalizada independientemente del valor que *Smooth* haya obtenido en la cata, y viceversa. En cambio, existe mayor probabilidad de que se valore la calidad del vino con un 5 si la suavidad del mismo es inferior al 2.75 aproximadamente.



**Figura 3.20:** Gráfica de efectos para la variable *Plante*.

Por lo contrario, aumenta la probabilidad de que un catador valore con un 6 la calidad del vino si se puntúa la suavidad con valores comprendidos entre el 1'45 y más del 4 aproximadamente, teniendo la probabilidad del 100 % entre los valores de suavidad 2.75 y 3.25. Por último, si se valora la suavidad con más de un 3.15 en puntuación, existe, y va aumentando, la probabilidad de obtener un 7 en la calidad general del vino. Lo que significa que, si se mejora la sensación de suavidad en el vino, se mejora la puntuación de calidad que se obtiene.



**Figura 3.21:** Gráfica de efectos para la variable *Smooth*.

## Capítulo 4

# Conclusiones

En este Trabajo Fin de Grado se ha estudiado el efecto que tienen las diferentes características de un vino percibidas en el proceso de la cata, con la calidad general que se aprecia del mismo, para así poder aplicar este conocimiento en empresas vitivinícolas y realizar mejoras en su proceso de producción.

En primer lugar, se ha hablado sobre la importancia del vino en la sociedad a lo largo de la historia, además de definir algunos de los componentes de la matriz del vino y de la percepción del mismo que resultan interesantes por su interacción con el producto final. Una vez se ha expuesto lo anteriormente mencionado, se ha argumentado la importancia que tiene el vino español a nivel nacional e internacional, por su aportación económica en indicadores tan importantes como el PIB o la repoblación rural, y en relación con la economía, se ha mencionado la relevancia de que las empresas conozcan en detalle dichos componentes del vino, cómo se comportan a lo largo del proceso de fabricación y cómo son percibidos por el consumidor, para que así se pueda determinar la matriz del vino que desencadene un producto de mejor calidad y más alineado con las preferencias de los consumidores.

En relación con los resultados obtenidos, cabe destacar lo siguiente:

Como bien se ha mencionado en el apartado 1.3, la determinación de los aromas en el vino es crucial para definir su calidad, y como se explica en Herderich y Smith 2005 el consumidor se centra mucho en los aromas a la hora de seleccionar un vino. Partiendo de esta información, se ha demostrado en el modelo estudiado que una mayor calidad del aroma percibido en el proceso de cata tiene una mayor probabilidad de que se puntúe mejor la calidad final del vino, existiendo de esta forma una relación positiva entre ambas variables. Esto se traduce en que las empresas vinícolas han de cuidar con especial precaución las levaduras y barricas que intervienen en el proceso de producción (Carpena et al. 2020, Carpena et al. 2020) porque si, por ejemplo, se transfieren al vino aromas como el de la madera, el producto va a poder competir en un mercado con un rango de precios más elevado (Juan et al. 2012). Las empresas tampoco han de descuidar la concentración tánica del vino, ni el estado de madurez con el que adquieren la uva, puesto que si la uva está poco madura y existe una alta concentración, el aroma puede ser eclipsado (Carpena et al. 2020).

Para cada una de las variables *Balance*, *Aroma.persistency*, *Attack.intensity*, *Plante* y *Smooth* del modelo estudiado, se ha obtenido una relación positiva con la variable respuesta *quality*. Todas estas variables son aspectos percibidos por los expertos catadores, y por tanto reflejan que son las características que los consumidores tienen en cuenta cuando toman un vino. Todas ellas son percibidas a través de los sentidos del gusto, olfato y vista, aunque también pueden intervenir otros sentidos como el tacto para modificar las percepciones (Wang y Spence 2018). Por este motivo, las empresas vinícolas han de tener en cuenta los efectos de los diferentes componentes de la matriz de un vino, dado que, por ejemplo, la suavidad de un vino puede verse afectada durante el proceso de envejecimiento por los pigmentos (antocianinas) existentes en la piel de la uva (SANCHEZ MARTINEZ DE ILARDUYA 2010).

En el caso contrario, la variable estudiada *Acidity* presenta una relación negativa con la variable respuesta *quality*. Los expertos catadores coinciden con los resultados obtenidos en el modelo, puesto que la percepción de la acidez del vino, viene determinada por la cantidad de ácido volátil que contiene el mismo, y dicho componente es el principal para la elaboración del vinagre (Gama et al. 2009).

Por último, en este trabajo, las variables *Density*, *Phenolic* y *chlorides* no tienen un efecto significativo sobre la calidad general percibida.

Se han realizado estudios similares al expuesto en este Trabajo Fin de Grado, donde se estudian las relaciones existentes entre ciertas variables y una variable respuesta en común entre todos ellos, que es la calidad general del vino. Estudios como Gama et al. 2009 y Valero y Cortijo 2013 utilizan la regresión lineal para generar su solución, o también la regresión hedónica como es el caso de Unwin 1999. En cambio, para la realización de este trabajo se ha utilizado como método la regresión ordinal dado que, se ha considerado más idóneo por las características de la variable respuesta. Por este motivo, y al aplicar un buen modelo, los parámetros obtenidos se ajustan mejor a los datos que otros modelos.

Una empresa vinícola conocedora de estas relaciones entre las diferentes variables es capaz de evitar costosos experimentos y pruebas de calidad del producto (Galeano-Arias, Aguirre y Castrillón-Gómez 2021), además de reducir el impacto de posibles pérdidas económicas por lotes fallidos causados por la calidad insuficiente del vino, lo que también se traduce en reducir la cantidad de desechos generados y de la pérdida de materia prima, en este caso la uva. Esto se relaciona con el Objetivo de Desarrollo Sostenible 12 “Producción y consumo responsable”, más concretamente con los objetivos 12.5 y 12.3, que hablan de la reducción de desechos y de la pérdida de alimentos respectivamente.

Ser consciente de estas relaciones y aplicarlas en el proceso productivo también ayuda tanto a las empresas emergentes como a las ya asentadas en el mercado, a ofrecer al consumidor un producto de mejor calidad y que se adapta de mejor forma a sus requisitos y gustos, desencadenando de esta forma un aumento en la satisfacción del consumidor y con ello, un posible aumento en las ventas de vino.

El sector vinícola se caracteriza por estar dividido en pequeñas plantaciones por toda España, y con ello es capaz de aportar al PIB nacional cerca de un 2% de su valor (Internacionales 2023), por lo que un aumento en las ventas provocará un incremento en este porcentaje, y un desarrollo y crecimiento de las pequeñas y medianas empresas que lo elaboran. De esta forma, el sector ayuda a que España cumpla las metas 8.1, 8.3 y 9.1, correspondientes al Objetivo de Desarrollo Sostenible “Trabajo decente y crecimiento económico” (ODS 8) e “Industria, innovación e infraestructura” (ODS 9).

---

Al ser un sector que se desarrolla en el medio rural, los trabajos que se generan fomentan que las personas residan en zonas rurales que estaban siendo despobladas, lo que desencadena que se vayan reduciendo las desigualdades existentes entre las ciudades y las zonas rurales y las personas encuentren trabajo digno sin tener que abandonar España (Sánchez-Hernández 2014). Esto se relaciona con los ODS 8 “Trabajo decente y crecimiento económico” y 11 “Ciudades y comunidades sostenibles”, más concretamente con las metas 11.a y de nuevo al 8.3. Además, de las 930.000 hectáreas de viñedos que se encuentran en España, 142.100 son ecológicas, lo que favorece a que se cumpla la meta 12.2 “De aquí a 2030, lograr la gestión sostenible y el uso eficiente de los recursos naturales”, perteneciente al ODS 12 “Producción y consumo responsable”.





# Bibliografía

- Alcalde-Eon, Cristina et al. (2014). «Adding oenological tannin vs. overripe grapes: Effect on the phenolic composition of red wines». En: *Journal of Food Composition and Analysis* 34.1, págs. 99-113 (vid. pág. 9).
- Belda, Ignacio et al. (2017). «Microbial contribution to wine aroma and its intended use for wine quality improvement». En: *Molecules* 22.2, pág. 189 (vid. pág. 7).
- Carpena, Maria et al. (2020). «Secondary aroma: Influence of wine microorganisms in their aroma profile». En: *Foods* 10.1, pág. 51 (vid. págs. 6, 7, 9, 33).
- Castaldo, Luigi et al. (2019). «Red wine consumption and cardiovascular health». En: *Molecules* 24.19, pág. 3626 (vid. pág. 4).
- Christensen, Rune H. B. (2023). *ordinal—Regression Models for Ordinal Data*. R package version 2023.12-4 (vid. pág. 15).
- Csárdi, Gábor et al. (2024). *igraph: Network Analysis and Visualization in R*. R package version 2.0.3. DOI: 10.5281/zenodo.7682609 (vid. pág. 15).
- De la Torre Pigazos, Marta (2019). «Exportación de vino tinto al Estado de California: un estudio de caso». En: *Studia Zamorensia* 18, págs. 235-255 (vid. pág. 2).
- Dunn, Peter K, Gordon K Smyth et al. (2018). *Generalized linear models with examples in R*. Vol. 53. Springer (vid. pág. 17).
- Fehér, János, Gabriella Lengyel y Andrea Lugasi (2007). «The cultural history of wine-theoretical background to wine therapy». En: *Central European Journal of Medicine* 2, págs. 379-391 (vid. pág. 1).
- Ferreira, Vicente, Arancha De-La-Fuente-Blanco y María-Pilar Sáenz-Navajas (2021). «A new classification of perceptual interactions between odorants to interpret complex aroma systems. Application to model wine aroma». En: *Foods* 10.7, pág. 1627 (vid. pág. 6).
- Fruchterman, Thomas M. J. y Edward M. Reingold (1991). «Graph Drawing by Force-directed Placement». En: *Software: Practice and Experience* 21.11, págs. 1129-1164 (vid. pág. 24).

- G Pereira, Antia et al. (2020). «Chapter Management of Wine Aroma Compounds: Principal Basis and Future Perspectives». En: (vid. pág. 6).
- Galeano-Arias, Luisa F, Sergio G Aguirre y Omar D Castrillón-Gómez (2021). «Análisis de calidad del vino por medio de técnicas de inteligencia artificial». En: *Información tecnológica* 32.1, págs. 17-26 (vid. pág. 34).
- Gama, João et al. (2009). *Discovery Science: 12th International Conference, DS 2009, Porto, Portugal, October 3-5, 2009*. Vol. 5808. Springer Science & Business Media (vid. pág. 34).
- Gambutì, Angelita et al. (2017). «Oxygen exposure of tannins-rich red wines during bottle aging. Influence on phenolics and color, astringency markers and sensory attributes». En: *European Food Research and Technology* 243, págs. 669-680 (vid. pág. 10).
- George Henry Dunteman, Dr. Moon-Ho R. Ho (2005). *An Introduction to Generalized Linear Models (Quantitative Applications in the Social Sciences)*. 1.<sup>a</sup> ed. Quantitative Applications in the Social Sciences 145. Sage Publications, Inc. ISBN: 0761920846; 9780761920847 (vid. págs. 16, 17).
- Herderich, MJ y PA Smith (2005). «Analysis of grape and wine tannins: Methods, applications and challenges». En: *Australian Journal of Grape and Wine Research* 11.2, págs. 205-214 (vid. pág. 33).
- Hernández Orte, Purificación (s.f.). «Influencia de la adición de taninos enológicos en el color del vino». En: () (vid. pág. 9).
- Intensos, Ruca Malén y Norberto Richardi (s.f.). «El efecto de la matriz del vino en la extracción de taninos durante la maceración y la percepción de astringencia pág 22». En: () (vid. pág. 6).
- Internacionales, Analistas Financieros (2023). *La relevancia económica y social del sector vitivinícola en España, 2023* (vid. págs. 2, 3, 34).
- Jackson, Ronald S (2008). *Wine science: principles and applications*. Academic press (vid. pág. 1).
- Jeromel, Ana, Ana-Marija Jagatić Korenika e Ivana Tomaz (2019). «An influence of different yeast species on wine aroma composition». En: *Fermented beverages*. Elsevier, págs. 171-285 (vid. pág. 7).
- Jordão, António M, Alice Vilela y Fernanda Cosme (2015). «From sugar of grape to alcohol of wine: Sensorial impact of alcohol in wine». En: *Beverages* 1.4, págs. 292-310 (vid. págs. 6, 7, 9).
- Juan, Felipe San et al. (2012). «Aroma chemical composition of red wines from different price categories and its relationship to quality». En: *Journal of Agricultural and Food Chemistry* 60.20, págs. 5045-5056 (vid. págs. 6, 33).
- Kuhn, Max, Simon Jackson y Jorge Cimentada (2022). *corr: Correlations in R*. R package version 0.4.4 (vid. pág. 14).

- 
- La rueda de los aromas del vino* (s.f.). <https://aromaster.com/es/product/rueda-de-los-aromas-del-vino/>. Accessed: 2024-04-03 (vid. pág. 8).
- Lachman, Jaromír et al. (2009). «Major factors influencing antioxidant contents and antioxidant activity in grapes and wines». En: *International Journal of wine research*, págs. 101-121 (vid. págs. 4, 6).
- Lipsitz, Stuart R et al. (2013). «Bias correction for the proportional odds logistic regression model with application to a study of surgical complications». En: *Journal of the Royal Statistical Society Series C: Applied Statistics* 62.2, págs. 233-250 (vid. pág. 27).
- Lüdecke, Daniel (2018). «ggeffects: Tidy Data Frames of Marginal Effects from Regression Models.» En: *Journal of Open Source Software* 3.26, pág. 772. DOI: 10.21105/joss.00772 (vid. pág. 14).
- Marin, David Hervas (2021). *repmod: Create Report Table from Different Objects*. R package version 0.1.7 (vid. pág. 15).
- (2023). *clickR: Semi-Automatic Preprocessing of Messy Data with Change Tracking for Dataset Cleaning*. R package version 0.9.39 (vid. pág. 14).
- Martín, David (jul. de 2017). «Desalcoholización y Recuperación de Aromas de Vino mediante Extracción con CO<sub>2</sub> Supercrítico». Tesis doct. DOI: 10.13140/RG.2.2.21152.53764 (vid. pág. 7).
- Martínez Montero, María et al. (2020). «STUDY OF THE PRODUCTION SYSTEM OF RED AND WHITE WINE AND PROPOSALS FOR IMPROVEMENT IN A WINE COMPANY». En: (vid. pág. 10).
- McRae, Jacqui M, Martin P Day et al. (2015). «Effect of early oxygen exposure on red wine colour and tannins». En: *Tetrahedron* 71.20, págs. 3131-3137 (vid. pág. 9).
- McRae, Jacqui M y James A Kennedy (2011). «Wine and grape tannin interactions with salivary proteins and their impact on astringency: a review of current research». En: *Molecules* 16.3, págs. 2348-2364 (vid. pág. 4).
- Muñoz, David et al. (2006). «Higher alcohols concentration and its relation with the biological aging evolution». En: *European Food Research and Technology* 222, págs. 629-635 (vid. pág. 6).
- Muñoz-González, Carolina, María Pérez-Jiménez y María Ángeles Pozo-Bayón (2020). «Oral persistence of esters is affected by wine matrix composition». En: *Food research international* 135, pág. 109286 (vid. pág. 8).
- NEWS, ARCHAEO (s.f.). *8,000-year-old wine unearthed in Georgia*. <https://www.stonepages.com/news/archaeo>. Accessed: 2024-06-05 (vid. pág. 1).
- Otamendi, Juan José (s.f.). «EL SERVICIO DEL VINO EL SERVICIO DEL VINO». En: () (vid. pág. 9).

- Padilla, Beatriz, José V Gil y Paloma Manzanares (2016). «Past and future of non-Saccharomyces yeasts: From spoilage microorganisms to biotechnological tools for improving wine aroma complexity». En: *Frontiers in microbiology* 7, pág. 185047 (vid. pág. 9).
- Palacios, Antonio y Nathalie Beaucourt (s.f.). «EFECTO EN BARRICAS DE 3 Y 4 AÑOS REGENARADAS MEDIANTE EL MÉTODO BARENA® SOBRE LA CESIÓN DE TANNINOS Y AROMAS YA NIVEL MICROBIOLÓGICO». En: *cresol* 1.1 (), pág. 1 (vid. pág. 10).
- Parry, Stephen (2010). «Ordinal Logistic Regression models and Statistical Software: What You Need to Know». En: págs. 1-4 (vid. pág. 16).
- Pedersen, Thomas Lin (2024). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. R package version 2.2.1 (vid. pág. 15).
- Perestrelo, Rosa et al. (2020). «An approach of the madeira wine chemistry». En: *Beverages* 6.1, pág. 12 (vid. pág. 7).
- Perez-Jiménez, María, Carolina Chaya y Maria Ángeles Pozo-Bayón (2019). «Individual differences and effect of phenolic compounds in the immediate and prolonged in-mouth aroma release and retronasal aroma intensity during wine tasting». En: *Food Chemistry* 285, págs. 147-155 (vid. pág. 8).
- Portela, Julio Fernández (2013). «La evolución reciente del sector vitivinícola internacional». En: *GeoGraphos: Revista Digital para Estudiantes de Geografía y Ciencias Sociales* 4.39, págs. 171-192 (vid. pág. 2).
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria (vid. págs. 11, 14, 15, 19).
- Redondo, Noemí, Sonia Gómez-Martínez y Ascensión Marcos (2014). «Sensory attributes of soft drinks and their influence on consumers' preferences». En: *Food & function* 5.8, págs. 1686-1694 (vid. pág. 7).
- RStudio Team (2023). *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA (vid. pág. 14).
- SANCHEZ MARTINEZ DE ILARDUYA, M<sup>a</sup> BELEN (2010). «Pigmentos derivados antociánicos de los vinos tintos de Rioja. Estudio analítico, influencia en el color y evolución durante la crianza». Tesis doct. Universidad del País Vasco-Euskal Herriko Unibertsitatea (vid. pág. 34).
- Sánchez-Hernández, José Luis (dic. de 2014). «El valor social y territorial del vino en España». En: págs. 31-66. ISBN: 978-84-95531-65-0 (vid. págs. 2, 35).
- Smith, PA, JM McRae y KA Bindon (2015). «Impact of winemaking practices on the concentration and composition of tannins in red wine». En: *Australian Journal of Grape and Wine Research* 21, págs. 601-614 (vid. págs. 4, 5).
- Suárez, Sabrina Rodríguez et al. (2018). «El vino, ¿beneficioso o perjudicial para la salud?». En: *Medicentro* 22.4, págs. 343-350 (vid. pág. 4).

- Unwin, Tim (1999). «Hedonic price indexes and the qualities of wines». En: *Journal of Wine Research* 10.2, págs. 95-104 (vid. pág. 34).
- Valero, Juan Sebastián Castillo y María Carmen García Cortijo (2013). «Analysis of international competitive positioning of quality wine from Spain». En: *Ciencia e investigación agraria: revista latinoamericana de ciencias de la agricultura* 40.3, págs. 491-501 (vid. pág. 34).
- Versari, Andrea, W Du Toit y GIUSEPPINA PAOLA Parpinello (2013). «Oenological tannins: A review». En: *Australian Journal of Grape and Wine Research* 19.1, págs. 1-10 (vid. pág. 5).
- Villamor, Remedios R y Carolyn F Ross (2013). «Wine matrix compounds affect perception of wine aromas». En: *Annual review of food science and technology* 4, págs. 1-20 (vid. pág. 3).
- Wang, Qian Janice y Charles Spence (2018). «A smooth wine? Haptic influences on wine evaluation». En: *International journal of gastronomy and food science* 14, págs. 9-13 (vid. pág. 34).
- Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4 (vid. pág. 14).
- Wilkinson, Leland (2012). *The grammar of graphics*. Springer (vid. pág. 14).