

Taller de práctica 9

Métodos de selección de variables, e intervalos de confianza y predicción
(secciones 3.4 y 3.5)

Los dos conjuntos de datos están relacionados con las variantes tinto ([link](#)) y blanco ([link](#)) del vino portugués “Vinho Verde”. Debido a cuestiones de privacidad y logística, sólo están disponibles variables fisicoquímicas (entradas) y sensoriales (salidas) (por ejemplo, no hay datos sobre tipos de uva, marca de vino, precio de venta del vino, etc.).

Las columnas de las bases de datos son las siguientes:

- **fixed.acidity** (g/L): cantidad de la mayoría de los ácidos involucrados con el vino o fijos o no volátiles (no se evaporan fácilmente).
- **volatile.acidity** (g/L): cantidad de ácido acético en el vino, que en niveles demasiado altos puede provocar un sabor desagradable a vinagre
- **citric.acid** (g/L): cantidad de ácido cítrico, que se encuentra en pequeñas cantidades, el cual, puede añadir “frescura” y sabor a los vinos.
- **residual.sugar** (g/L): cantidad de azúcar que queda después de que se detiene la fermentación. Es raro encontrar vinos con menos de 1 gramo/litro y los vinos con más de 45 gramos/litro se consideran dulces.
- **chlorides** (g/L): cantidad de sal en el vino.
- **free.sulfu.dioxide** (ppm): cantidad de la forma libre de SO_2 existe en equilibrio entre el SO_2 molecular (como gas disuelto) y el ion bisulfito; previene el crecimiento microbiano y la oxidación del vino.
- **total.sulfur.dioxide** (ppm): cantidad de formas libres y ligadas de SO_2 ; En concentraciones bajas, el SO_2 es prácticamente indetectable en el vino, pero en concentraciones de SO_2 libre superiores a 50 ppm, el SO_2 se vuelve evidente en la nariz y el sabor del vino.
- **density** (g/cm^3): densidad del agua dependiendo del porcentaje de alcohol y contenido de azúcar.
- **pH**: describe qué tan ácido o básico es un vino en una escala de 0 (muy ácido) a 14 (muy básico); la mayoría de los vinos tienen entre 3 y 4 en la escala de pH.
- **sulphates** (mg/L): cantidad de aditivo del vino que puede contribuir a los niveles de dióxido de azufre (SO_2), que actúa como antimicrobiano y antioxidante.
- **alcohol** (ABV, Alcohol por volumen): el porcentaje de alcohol del vino.
- **quality**: variable de salida (basada en datos sensoriales, puntuación entre 0 y 10), un número cercano a 10 indica un vino de mayor calidad.

A continuación, se utiliza la base de datos asociada a los datos del vino tinto para los enunciados siguientes.

1. Ajuste un modelo para estudiar el nivel de pH del vino a través de las variables **chlorides**, **density** y **sulphates**.
2. Ajuste un modelo para estudiar el nivel del pH, utilizando la metodología *backward*.
3. ¿Es posible comparar el desempeño de los dos modelos ajustados anteriormente? En caso de ser posible, compare e interprete.
4. Ajuste un tercer modelo para estudiar el nivel del pH, utilizando la metodología *forward*. Comente similitudes y diferencias respecto al modelo ajustado con la metodología *backward* con relación a las variables utilizadas por cada uno de los modelos.
5. Considerando el modelo ajustado en el ítem 1, responda.
 - a) ¿Cuál sería el nivel de pH de un vino con una cantidad de sal de $0.05 g/L$, una densidad del agua de $0.99 g/cm^3$ y una cantidad de aditivo del vino que puede contribuir a los niveles de dióxido de azufre de $0.63 mg/L$?

- b) ¿Cuál sería el nivel de pH promedio de un vino con una cantidad de sal de 0.05 g/L , una densidad del agua de 0.99 g/cm^3 y una cantidad de aditivo del vino que puede contribuir a los niveles de dióxido de azufre de 0.63 mg/L ?
- c) Comente las razones de la diferencias entre los intervalos de confianza obtenidos en los ítems anteriores.