

# Modelado estadístico de la producción de vinos tintos finos

## Statistical modelling of fine red wine production

María Rosa Castro<sup>1</sup>, Marcelo Eduardo Echeagaray<sup>2</sup>, Rosa Ana Rodríguez<sup>3</sup> y Stella Maris Udaquiola<sup>4</sup>

### RESUMEN

En la provincia de San Juan, Argentina, la elaboración del vino es una de las actividades económicas más importantes. Para este tipo de industria, es de fundamental importancia predecir la cantidad de producción en función de la cantidad de materia prima. El objetivo de este trabajo es la obtención de un modelo que relacione los kilogramos de uva molidos con los litros de vino producidos. Dicho modelo se usará para realizar predicciones de valores futuros, puntuales y por intervalos de confianza para un nivel determinado de molienda. A tal fin, se trabajó con datos obtenidos de una bodega ubicada en la provincia de San Juan. En primer lugar, se calculó el coeficiente de correlación muestral y se realizó un diagrama de dispersión, el cual indicó una marcada relación lineal entre los litros de vino obtenidos y los kilogramos de uva molidos. Posteriormente, se adoptaron dos modelos lineales y se efectuó el análisis de varianza, ya que los datos provenían de poblaciones normales con igual varianza. De este análisis se desprendió el modelo más adecuado, el cual fue validado con valores experimentales, obteniendo una buena aproximación de estos.

**Palabras clave:** industria vitivinícola, modelo probabilístico, regresión lineal.

### ABSTRACT

Producing wine is a very important economic activity in the province of San Juan in Argentina; it is therefore most important to predict production regarding the quantity of raw material needed. This work was aimed at obtaining a model relating kilograms of crushed grape to the litres of wine so produced. Such model will be used for predicting precise future values and confidence intervals for determined quantities of crushed grapes. Data from a vineyard in the province of San Juan was thus used in this work. The sampling coefficient of correlation was calculated and a dispersion diagram was then constructed; this indicated a lineal relationship between the litres of wine obtained and the kilograms of crushed grape. Two lineal models were then adopted and variance analysis was carried out because the data came from normal populations having the same variance. The most appropriate model was obtained from this analysis; it was validated with experimental values, a good approach being obtained.

**Keywords:** wine industry, probabilistic model, lineal regression.

**Recibido:** junio 6 de 2008

**Aceptado:** septiembre 14 de 2009

### Introducción

En la provincia de San Juan, Argentina, la elaboración del vino es una de las actividades económicas más importantes. El proceso de producción del vino varía enormemente de acuerdo al lugar, la bodega, y la calidad y variedad de vino. Mientras el proceso usual puede ser resumido a una forma simplificada, hay muchas diferencias en este proceso que pueden afectar el vino obtenido, como corrientes requeridas, residuos, tiempo de almacenamiento, energía, requerimientos de limpieza y efectos ambientales asociados. El proceso de vinificación puede dividirse en las operaciones unitarias que se llevan a cabo durante él. La vendimia es el período en el que la bodega presenta su máximo funcionamiento, llevándose a cabo la cosecha y la molienda de la uva (Chapman, *et ál.* 2001).

El proceso de elaboración del vino comienza con la recolección de la uva. Ésta se realiza cuando el fruto se encuentra en el estado

óptimo de maduración, ya que condiciona la calidad e incluso el tipo de vino. Es muy importante el estado sanitario de la uva y su integridad para evitar fermentaciones prematuras e intercambios entre mosto y raspón, que originarían aromas y gustos no deseados.

El transporte a la bodega se hace en remolques de fácil volcado o en cajas en el menor tiempo posible. El remolque debe ser de acero inoxidable, estar protegido con una lona y provisto de un doble fondo para evitar la maceración del mosto. Las cajas, de una capacidad aproximada de 20 kg de uva cada una, se transportan en el remolque y se descargan manualmente en la bodega. Este sistema evita que la uva se aplaste por su propio peso y que llegue por lo tanto en mejores condiciones a la bodega (Storm, 2001).

Las uvas se descargan en una tolva de acero inoxidable provisto en el fondo de un tornillo sinfín que las conduce hacia la prensa, estrujador o estrujadora - despalilladora.

<sup>1</sup> Ingeniera química. Jefe de Trabajos Prácticos, Instituto de Ingeniería Química, Facultad de Ingeniería, Universidad Nacional de San Juan, Argentina. mrcastro@unsj.edu.ar

<sup>2</sup> Ingeniero químico. Jefe de Trabajos Prácticos, Instituto de Ingeniería Química, Facultad de Ingeniería, Universidad Nacional de San Juan, Argentina. mechega@unsj.edu.ar

<sup>3</sup> Ingeniera química. M.Sc., en Técnica Ambiental. Ph.D., en Ingeniería. Jefe de Trabajos Prácticos, Instituto de Ingeniería Química, Facultad de Ingeniería, Universidad Nacional de San Juan, Argentina. rrodri@unsj.edu.ar

<sup>4</sup> Ph.D., en Ingeniería Química. Profesor Titular Exclusivo, Instituto de Ingeniería Química, Facultad de Ingeniería, Universidad Nacional de San Juan, Argentina. estelaudaquiola@unsj.edu.ar

Con el prensado se extrae el mosto de la uva fresca. Esta operación determinará la calidad del vino. Se persigue obtener el mayor mosto posible aplicando la mínima presión para evitar la rotura de pepitas y raspones que transmitirían sustancias de sabor desagradable al mosto. El prensado se debe ejecutar en el menor tiempo posible para reducir la incorporación de aire.

Es muy común realizar un estrujado antes del prensado para facilitar la extracción del mosto. El estrujado consiste en reventar la uva para liberar el jugo de las células de la pulpa haciéndolas pasar entre dos cilindros muy próximos entre sí que giran en sentido contrario. La distancia entre los dos rodillos debe causar la rotura de las pepitas. Una vez terminado el estrujado de la uva en las máquinas correspondientes, existe una cantidad de mosto libre que se puede separar por simple decantación o mediante máquinas escurridoras. El estrujado debe ser hecho de manera rápida con el fin de limitar la intensidad de los fenómenos de maceración y de oxidación. Después del escurrido, la uva recolectada pasa a las prensas para la separación del mosto que aún le queda. En el caso de no estrujar la uva, se hace un prensado directo, cargando la prensa con uvas enteras.

Posteriormente al prensado, se realiza el sulfitado, que consiste en la adición de dióxido de azufre (SO<sub>2</sub>). Esta sustancia se agrega para controlar los microorganismos evitando que la fermentación se produzca en forma tumultuosa, inhibiendo el crecimiento de las levaduras no productoras de alcohol y favoreciendo las productoras de él. Además protege el mosto de la oxidación por el oxígeno (O<sub>2</sub>) del aire, ya que éste desnaturaliza el aroma, destruye el afrutado y oscurece el vino blanco; también ayuda a la clarificación del mosto debido a que el retardo de la fermentación implica que durante este periodo el mosto puede decantar sus materias en suspensión (Oreglia, 1978).

Luego, el mosto obtenido, conteniendo hollejos y pulpa, se vuelca en grandes depósitos de madera o acero inoxidable, donde va fermentando debido a las levaduras, transformando los azúcares en alcohol etílico y otros elementos, además de desprender anhídrido carbónico. A su vez, al encontrarse el mosto en contacto con las partes sólidas de la uva, va extrayendo de éstas el color de los denominados antocianos, taninos, etcétera.

El gas carbónico resultante empuja hacia arriba los hollejos, formando en la parte superior una capa llamada sombrero, la cual se debe ir remojando para que continúe la extracción de color y otros elementos. Para ello se realiza el remontado, consistente en extraer líquido de la parte inferior del depósito mediante una manguera e introducirlo por la parte superior. Además, para que el sombrero no se haga excesivamente compacto, debe ser removido cada cierto tiempo. A esto último se le llama bazuqueo (Storm, 2001).

En función del tipo de vino que se quiera elaborar se hará este proceso con más o menos tiempo, habitualmente de 8 y 12 días, a una temperatura entre 26 a 29°C (Santamaría *et ál.*, 1995; Jolibert, 1991). Posteriormente se realiza el descube, que consiste en trasegar únicamente el líquido a otro depósito. La parte sólida restante está impregnada de mosto y para extraerlo se utilizan prensas que aprietan estas masas hasta casi secarlas. El líquido resultante, denominado vino de prensa, es mucho más rico en color y taninos y no se mezcla con el anterior.

Tras la primera fermentación, se lleva a cabo la fermentación maloláctica, el paso de ácido málico a ácido láctico. Este líquido es mucho más suave y agradable que el anterior, por cuanto el vino gana en finura (Ziraldó, 2002).

En los vinos nuevos se produce una clarificación espontánea, depositando en el fondo de las cubas "las madres" (lías, fangos). Es aconsejable que estos sedimentos no estén mucho tiempo junto al vino para ir disminuyendo la turbidez. Por esta razón se trasega el vino a cubas limpias frecuentemente. Este proceso airea el vino, siendo esto conveniente al principio, para ayudar al buen acabado de la fermentación y la estabilización del vino, permitiendo la evaporación de sustancias volátiles resultantes de la fermentación y de gas carbónico. El trasego también es usado para la homogeneización de vinos entre diferentes cubas con el fin de conseguir uniformidad.

Aunque en el trasegado muchos elementos en suspensión son retirados del vino, otros más ligeros no llegan a decantar por sí solos. Por ello se agregan al vino sustancias coloides que arrastran hacia el fondo impurezas en suspensión (Yair, 1996).

Con la filtración se separa la fase sólida insoluble de la fase líquida, sin modificar las características del vino. Durante la elaboración de éste se van a realizar tres filtraciones, con la misma finalidad de limpieza, pero con diferencia en el desarrollo, características y maquinaria empleada. Las dos primeras, filtración devastadora y filtración abrillantadora, se hacen con un solo filtro de tierras, mientras que la tercera, o filtración amicróbica, con filtro de membrana (Datz y Kullmann, 2006).

La utilización de frío en la bodega se centra en dos puntos principales: control de la temperatura de fermentación y estabilización del vino por ultrarrefrigeración (Palacios *et ál.*, 2009).

El proceso de embotellado constará de varias etapas: despaletizado de las botellas, lavado, llenado, taponado, capsulado, etiquetado, encartonado y embalado de las botellas en cajas, las cuales se organizan en *pallets* y se almacenan hasta su expedición.

Los vinos jóvenes son comercializados una vez embotellados; sin embargo, los vinos de crianza y reserva reposan en barricas de madera antes de ser embotellados y etiquetados.

Uno de los desafíos que se presenta en la producción en general, es el de conocer la relación existente entre la cantidad de materia prima utilizada y la cantidad de producto obtenido. Para ello se deben utilizar técnicas estadísticas tales como el análisis de regresión, que nos permitirá explorar la naturaleza de esta relación y obtenerla (Flanzy, 2003).

Teniendo en cuenta lo expuesto es que se planteó como objetivo de este trabajo la modelización e investigación de la relación existente entre dos variables muy importantes que se presentan en el proceso de vinificación: los kilogramos de uva molida frente a litros obtenidos de vino tinto fino. Se trabajó con los datos de cantidades de uva molida y litros de vinos obtenidos, proporcionados por una bodega de gran envergadura ubicada en la provincia de San Juan. Para el procesamiento estadístico de ellos se utilizó el software Statgraphics (Pérez López, 1998). En primer lugar, se calculó el coeficiente de correlación muestral que mide la fuerza de la asociación lineal de las variables y se hizo un diagrama de dispersión, el cual indicó una marcada relación lineal entre los litros obtenidos de vino y kilogramos de uva molidos. Posteriormente, se adoptaron dos modelos lineales y se realizó el análisis de varianza, ya que los datos provenían de poblaciones normales con igual varianza. De este análisis se desprendió el modelo más adecuado, el cual fue validado con valores experimentales.

### Análisis estadístico de los datos

Para llevar a cabo este trabajo, se cuenta con los datos reales de kilogramos de uva molida y litros de vino fino obtenidos en una bodega de nuestro medio, entre los años 1994 a 2002.

Los datos proporcionados por la bodega se presentan en la tabla 1. Éstos fueron obtenidos siguiendo el proceso productivo descrito con anterioridad. Para el procesamiento de ellos se optó por el software Statgraphics (Pérez López, 1998), una herramienta utilizada para el procesamiento estadístico de datos ya que permite el análisis de los resultados, su interpretación y la elaboración de conclusiones.

Tabla 1. Datos proporcionados por la bodega testigo

Año	Cantidad de uva molida (kg)	Cantidad de vino obtenido (l)
1994	4.811.07	3.918.00
1995	4.729.480	3.761.595
1996	4.446.230	3.510.000
1997	5.487.166	4.320.000
1998	6.529.110	5.140.220
1999	6.762.010	5.370.000
2000	11.699.650	8.947.060
2001	11.730.510	8.956.650
2002	8.754.460	6.560.133

Una estadística descriptiva de los datos sobre kilogramos de uva molidos frente a litros de vino obtenidos, arrojó los resultados desplegados en la tabla 2.

Se puede verificar la suposición de normalidad de las poblaciones comparadas controlando que los coeficientes de forma y apuntamiento se encuentren dentro del rango de valores aceptables (-2,2). Del análisis de los coeficientes estandarizados de Skewness y Kurtosis surge el poder considerarse que los datos provienen de poblaciones normales (Box et ál., 2005).

Tabla 2. Datos descriptivos de las cantidades de materia prima utilizada y producto obtenido.

	Cantidad de uva molida (kg)	Cantidad de vino obtenido (l)
Media	7,21663E6	5,6093E6
Varianza	8,26946E12	4,48383E12
Desviación estándar	2,87567E6	2,11751E6
Coef. Skewness	1,07226	1,09655
Coef. Kurtosis	-0,485889	-0,412561

### Descripción gráfica de los datos numéricos

En la figura 1 se ofrecen los diagramas de cajas y bigotes, lo cual da una idea visual sobre la variabilidad de los datos y al no existir algún dato sospechoso que se deba eliminar, para evitar la distorsión de las soluciones obtenidas.

La base de la caja del diagrama se extiende desde el primer cuartil hasta el tercero. La caja representa la mitad central de los datos. La barra vertical que la fracciona en dos partes corresponde a la mediana, y el signo + señala el promedio de los datos (media muestral).

Los bigotes abarcan los datos que apartan como máximo 1,5 veces la diferencia entre el tercer cuartil y el primero (rango intercuartílico). Si algún dato supera estos límites se considera sospechoso y se debería estudiar su procedencia. En este caso en particular como la caja correspondiente a kilogramos molidos es más ancha que la correspondiente a litros obtenidos, indica que los datos de kilogramos molidos están más dispersos que los de litros obtenidos, y co-

mo los bigotes del primer diagrama son más largos, que tienen mayor variabilidad.

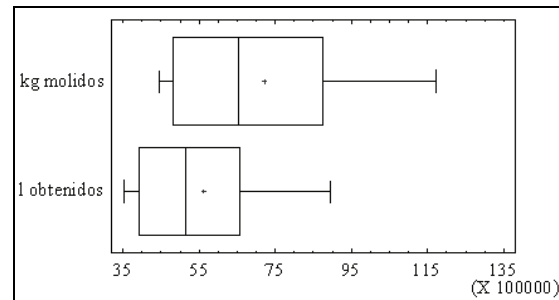


Figura 1. Diagrama de caja y bigotes

Además se puede observar en ambos gráfico que como la media está próxima a la mediana y ambas partes de la caja son aproximadamente iguales, los datos presentan simetría con respecto a un valor central.

### Coefficiente de correlación muestral

Se comienza el análisis de los datos calculando el coeficiente de correlación muestral, comprendido entre (-1,1) y mide la fuerza de la asociación lineal entre las variables. La tabla 3 despliega las correlaciones entre las variables. En primer lugar, se puede observar el coeficiente de correlación muestral. Entre paréntesis se indica la cantidad de datos analizados. Debajo de cada coeficiente figura el valor P de la prueba, que establece si la relación encontrada es estadísticamente significativa. Valores P por debajo de 0,05 indican correlación significativa.

Tabla 3. Correlaciones entre las variables

	Cantidad de uva molida (kg)	Cantidad de vino obtenido (l)
Cantidad de uva molida (kg)		0,9992 (9) 0,0000
Cantidad de vino obtenido (l)	0,9992 (9) 0,0000	

Con los datos originales se realiza un diagrama de dispersión (Figura 2); el examen de este diagrama indica que existe una marcada relación entre litros obtenidos y kilogramos molidos, y parece ser razonable la consideración tentativa del modelo lineal.

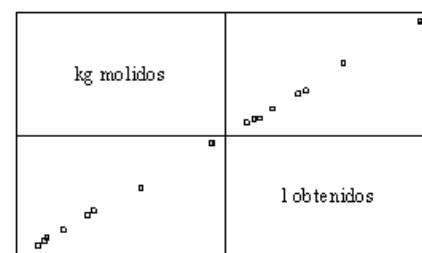


Figura 2. Matriz de dispersión

### Elección del modelo

Se ajusta en primer término al modelo lineal (Montgomery et ál., 2004):

$$y = \beta_0 + \beta_1x + \varepsilon \tag{1}$$

La salida de StatGraphics ofrece el siguiente modelo ajustado (Pérez López, 1998):

$$y = 299440,0 + 0,73578x \quad (2)$$

Sin embargo, a pesar de que el modelo en general es adecuado ya que el valor P de la prueba es 0 (Tabla 5), y por lo tanto es menor que el nivel de significación 0,05, el coeficiente de correlación es 0,999222, muy cercano a 1, y las pruebas de hipótesis:

$$\begin{aligned} H_0: \beta_0 = 0 & & H_0: \beta_1 = 0 \\ H_1: \beta_0 \neq 0 & & H_1: \beta_1 \neq 0 \end{aligned} \quad (3)$$

se rechazan (Tabla 4), es decir, a nivel poblacional la ordenada al origen y la pendiente de la recta de regresión son distintas de cero; el error estándar para la ordenada al origen no es del mismo orden que el error estándar de la pendiente, por lo tanto se decide ajustar a un modelo sin constante.

Tabla 4. Análisis de regresión para el modelo lineal

Parámetro	Estimación	Error estándar	Estadístico T	Valor P
Ordenada al origen	299440,0	84602,5	3,53937	0,0095
Pendiente	0,73578	0,0109743	67,0455	0,0000

Tabla 5. Análisis de varianza del modelo lineal

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	Razón F	Valor P
Modelo	3,58149E13	1	3,58149E13	4495,10	0,0000
Residuos	5,57728E10	7	7,96754E9		
Totales	3,58706E13	8			

El modelo sin constante tiene la forma (Montgomery et ál., 2004):

$$y = \beta x + \varepsilon \quad (4)$$

La salida computacional que muestra el modelo ajustado es la siguiente:

$$y = 0,772141x \quad (5)$$

El coeficiente de correlación muestral obtenido es igual a 0,999512.

El análisis de varianza para el ajuste al modelo sin constante arroja un valor P igual a cero, por lo cual se concluye que la regresión es significativa (Tabla 7).

Tabla 6. Análisis de regresión para el modelo lineal sin constante.

Parámetro	Estimación	Error estándar	Estadístico T	Valor P
Cant. de uva molida (kg)	0,772141	0,00602991	128,052	0,0000

Tabla 7. Análisis de varianza del modelo lineal sin constante.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	Razón F	Valor P
Modelo	3,18893E14	1	3,18893E14	16397,27	0,0000
Residuos	1,55583E11	8	1,94479E10		
Totales	3,19048E14	9			

La tabla 6 contiene la solución obtenida, las estadísticas T y los valores P para la prueba acerca de la significación del coeficiente de regresión  $\beta$ . Se observa que el valor P correspondiente a la prueba para  $\beta \neq 0$  es igual a cero, lo que conduce al rechazo de la hipótesis nula  $H_0: \beta = 0$  con un nivel de significación 0,05.

Un intervalo de confianza al 95% para el coeficiente estimado  $\beta$ , con un error máximo de estimación 0,00602991, es el que figura a continuación (0,758236, 0,786046).

## Discusión

Comparando los parámetros estadísticos obtenidos en el modelo representado por la ecuación (2) y el segundo modelo ensayado, representado por la ecuación (5), se concluye que el modelo estimado:

$$I_t \text{ obtenidos} = 0,772141 * \text{kg molidos}$$

representa apropiadamente los datos, y la figura 3 muestra la concordancia entre los valores observados y predichos.

Puesto que el modelo encontrado representa adecuadamente los datos, se puede usar para efectuar predicciones de valores futuros, puntuales y por intervalos de confianza.

## Validación del modelo adoptado

Con el objetivo de validar el modelo adoptado, se utilizó para predecir los litros de vino obtenidos en el año 2005 en función de los kilogramos de uva molidos.

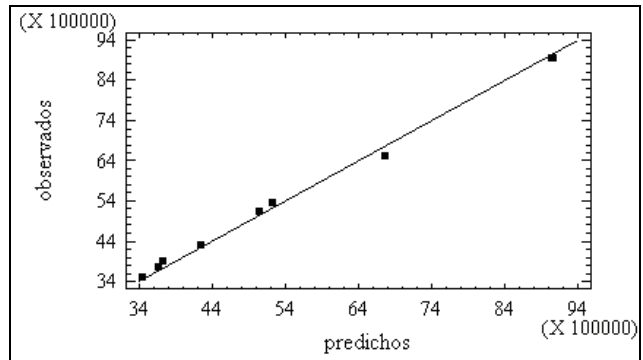


Figura 3. Modelo ajustado sin constante

Para 12.000.000 kg de uva molida, según el modelo lineal sin constante, se obtendrían 9.265.692 lt de vino. El intervalo de confianza a un 95% es (9.123.012, 9.408.372). Comparando el valor estimado con el valor real de 9.230.000 lt de vino, éste se encuentra contenido en él, lo que demuestra que el modelo al cual se arriba ajusta adecuadamente los datos.

En la figura 4 se puede apreciar un gráfico de probabilidad normal para los residuos. Puesto que la mayoría de ellos están cercanos a la recta, se puede considerar que provienen de una población normal. Una estadística descriptiva para los residuos proporciona un promedio muestral próximo a 37036,0 y desviación estándar igual a 133809,0; los coeficientes de Skewness y Kurtosis son -0,903952, -0,395964 respectivamente, y se encuentran en el intervalo (-2,2) (Box et ál., 2005).

## Conclusiones

El modelo lineal logrado, aunque sencillo permite observar la dependencia de los litros obtenidos en un proceso de vinificación con respecto a los kilogramos de uva que ingresan a una bodega, realizando total abstracción del resto de las variables que en un proceso de esta envergadura se presentan. Es útil para representar los datos y es una alternativa válida para modelar matemáticamente este tipo de procesos, en los cuales todavía el arte de la elaboración del vino y el criterio del enólogo son primordiales a la hora de la obtención del producto. El modelo que en este trabajo se presenta está validado con rigurosidad y por lo tanto es muy apropiado para realizar predicciones de valores futuros, puntuales y por

intervalos de confianza, siendo aplicable a bodegas de gran envergadura para la producción de vinos tintos finos, similar a la estudiada en este trabajo.

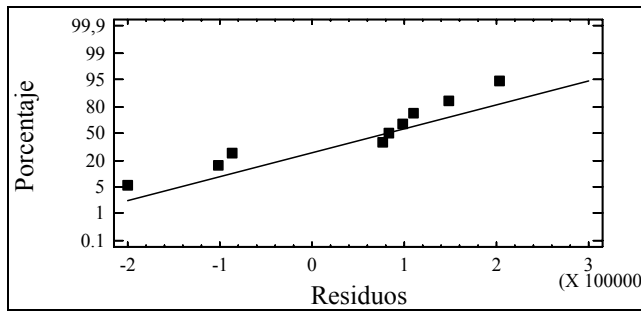


Figura 4. Gráfico de probabilidad normal para los residuos

### Nomenclatura

y: Litros de vino obtenidos

x: Kilogramos de uva molidos

$\beta, \beta_0, \beta_1$ : Coeficientes de los distintos modelos lineales.

$\epsilon$ : Error

$H_0, H_1$ : Hipótesis

### Bibliografía

Box, G., Stuart Hunter, W., Estadística para Investigadores. Introducción al diseño de experimentos y análisis de datos., Ed Reverté S.A., México, 2005.

Datz, C., Kullmann, C., Winery Design., The Neues Publishing Company, 2006.

Flanzy, C., Enología: Fundamentos Científicos y Tecnológicos., AMV Ediciones, 2003.

Jolibert, F., Consommations energetiques et filière vinicole., Bios, 5 (22), 1991, pp. 51-53.

Montgomery, D., Peck, E., Vining, G., Introducción al Análisis de Regresión Lineal., Ed. CECSA, México, 2004.

Oreglia, F., Enología Teórico-Práctica., Instituto Salesiano de Artes Gráficas, Buenos Aires, 1978.

Palacios, C., Udaquiola, S., Rodriguez, R., Modelo matemático para la predicción de las necesidades de frío durante la producción de vino., Ciencia, Docencia y Tecnología, Vol. 38, 2009.

Pérez López, C., Métodos Estadísticos con Statgraphics para Windows., Ed. RA-MA, 1998.

Santamaría, P., López, R., Gutierrez, A., García-Escudero E. Influencia de la temperatura en la fermentación alcohólica., Zúbia, No. 7, 1995, pp. 137-149.

Storm, D., Winery Utilities: Planning, Design and Operation., Ed. Kluwer Academia / Plenum Publisher, 2001.

Yair, M., Winery Technology & Operations: A Handbook for Small Wineries., Wine Appreciation Guild, 1996.

Ziraldo, D., Anatomy Of A Winery: The Art Of Wine At Inniskillin., Key Porter Books, 2002.