



Facultad de Ingeniería Mecánica

Título: Metodología para la construcción de modelos matemáticos de caja negra para la simulación dinámica de enfriadoras de agua condensadas por aire.

AUTOR: Gilberto José Francisco Pedro

Tutores: Msc. Yamile Díaz Torres

MSc. Miguel Santana Justiz

Cienfuegos. Cuba
Curso 2018 – 2019

“La precisión en el objetivo es el punto de partida para cualquier triunfo.”

W. Clement Stone

AGRADECIMIENTOS

Me es imposible en una cuartilla expresar mi agradecimiento a todas aquellas personas que de una forma u otra me ha ayudado a realizar mis sueños

En especial quiero agradecer a:

- *A Dios por todo.*
- *Mis padres, fuentes de amor y dedicación en mi vida como estudiante*
- *Mis tutores: Msc. Yamile Díaz Torres y MSc. Miguel Santana Justiz por poner a mi disposición su sabiduría, experiencia y ayuda en todo momento.*
- *A todos y cada uno de los profesores que con todo su esfuerzo influyeron en mi formación académica en el transcurso de mi carrera.*
- *A nuestra mami Lázara Pérez Clemente más conocida por mami Lachy que siempre estuvo presente en mi vida en el transcurso de mi carrera.*

DEDICATORIA

Dedico esta tesis a:

Mis padres, por la educación que me han dado, por todo su amor y esfuerzo

Resumen

La presente investigación tiene como objetivo presentar una metodología para analizar y evaluar modelos matemáticos de caja negra para la simulación energética de enfriadoras de agua helada condensadas por aire. Para el mismo se emplean el método de los mínimos cuadrados generalizado en la construcción modelos de regresión lineal. Se construyeron modelos lineales, cuadráticos, univariable y bivariantes. Estos fueron comparados empleando métricas como: el coeficiente de correlación de Pearson (R^2), el error cuadrático medio (RMS), Análisis de Varianza, Prueba t en la regresión, Análisis de los supuestos de la regresión.

Como resultado se obtuvo que, para los modelos de regresión lineal, se escogió el modelo de regresión lineal múltiple considerando el principio de parsimonia. Los resultados demuestran que son capaces de simular con mínimo error el funcionamiento de la enfriadora, lo cual permitirá en un futuro trazar estrategias de operación que incrementen la eficiencia del mismo.

Summary

Contenido

INTRODUCCIÓN.....	11
CAPITULO I Estado del arte del modelado energético de enfriadoras de agua....	14
1.1 Enfriadoras de agua. Fundamentos básicos.	14
1.2. Evaluación energética del funcionamiento de un chiller.....	16
1.2.1. Modelos de caja blanca	17
1.2.2. Modelos de caja negra.....	18
1.2.3 Modelos de caja gris.....	19
1.3 Modelos de caja negra basados en la regresión lineal en problemas	21
Conclusiones parciales	22
CAPITULO II Materiales y métodos	24
2.1 Fases de construcción de un modelo	25
2.1.1Selección del modelo de regresión	25
2.2 Obtener el modelo	28
2.2.1 Calcular los coeficientes	28
2.2.2 Análisis de varianza o prueba F en la regresión.....	29
2.2.3 Prueba t en la regresión	31
2.2.4 Análisis de los supuestos de la regresión	32
2.3 Conclusiones parciales.....	40
CAPITULO III Elaboración de modelos matemáticos para la simulación energética de enfriadora de agua helada a flujo constante.....	41
3.1 Selección de las variables.....	42
3.2 Selección del modelo de caja negra.....	43
3.3 Obtener el modelo	44
3.3.1 Variante uno, modelo de regresión lineal simple.	45
3.3.2 Variante dos del modelo de regresión lineal simple.....	46
3.3.3 Variante tres, modelo de regresión polinomial de orden 2.	50
3.3.4 Variante cuatro, modelo de regresión polinomial de orden 2.	50
3.3.5 Variante cinco, modelo de regresión lineal múltiple	55
3.3.6 Variante seis, modelo de regresión no lineal.....	59
3.3.7 Variante siete, modelo de regresión no lineal	63
3.3.8 Variante ocho del modelo no lineal	67

3.3.9 Variante nueve del modelo matemático	71
3.3.10 Variante diez del modelo matemático.....	74
3.4. Comparación de modelos	78
Conclusiones parciales	81
Conclusiones Generales	82
Recomendaciones	83
Bibliografía.....	84

INTRODUCCIÓN

Los sistemas HVAC son una premisa en el confort de la vida moderna, en los países tropicales las personas pasan más del 60% de su tiempo en lugares climatizados. Los sistemas HVAC son altos consumidores de energía, en un sistema centralizado las máquinas fundamentales que son los chiller alcanzan un 50% de la energía total del sistema y el sistema el 60% de toda instalación, por lo que el control de este sistema es vital para poder manejar correctamente los recursos energéticos.

Para efectuar el diseño de una planta se recomienda por varios autores(Dym, 2004)(Ardakani et al, 2008) que la misma debe partir del análisis de su futura explotación por lo que se requiere construir los modelos matemáticos para la simulación.El objetivo de la simulación del modelo consiste en no conocer el sistema en sí, sino el comportamiento ante diversas situaciones.Es importante la utilización de la simulación del modelo, aunque existen ecuaciones formuladas, pero resulta más sencillo y barato simular.

Existen tres enfoques genéricos del modelado matemático que han sido usados para la optimización de sistemas de climatización(Jiang, Framework Combining Static Optimization, 2005). Modelos de caja blanca, modelos de caja negra y modelos de caja gris. A pesar de las ventajas y desventajas de cada uno de ellos, los modelos de caja negra son los que más se emplean en la ingeniería por su simplicidad y precisión.La elaboración de un modelo de este tipo depende de los datos de funcionamiento de una enfriadora, para lo cual se necesita obtener los datos del fabricante o datos reales para su construcción y validación.

Algunos autores han presentado metodologías para la construcción y validación de modelos, por ejemplo, La Guía 14 de ASHRAE (2014) despliega una guía para desarrollar modelos de regresión para las edificaciones, así como un conjunto de métricas estadísticas para evaluar la calidad del modelo.(Wang, 2017) presentó una metodología para evaluar el consumo energético de chiller centrífugos. Gunay y autores 2017 emplearon la matriz de correlación para evaluar la multicolinealidad de las diversas variables que afectan el funcionamiento de los chillers. Por otro lado, es usual presentar

solamente un modelo de regresión y mostrar solamente valores de indicadores estadísticos como: el coeficiente de correlación $[R^2]$, RMSE, coeficiente de variación de los residuos, entre otros como es el caso de (Ho, 2019) (Jayasekara S. Halgamuge, 2014) (Seo, 2016), (Wang, 2017) (Ardakani et al, 2008); (Chang e. a., 2004) (Chang Y.-C. , 2007); (Santos et al, 2014) ; (Beghi et al, 2011); (Seo et al, 2016)

Por otro lado muchos estudios exponen el modelo matemático que describe el funcionamiento de una enfriadora por ejemplo (Ardakani, 2008); (Beghi, 2011); (Chang Y.-C. , 2004); (Chang, Lin, & Huang., 2005); (Chang Y.-C. , 2007); (Seo, 2016)). Pero los mismos no declaran como obtuvieron el modelo, o expusieron distintos modelos y compararon la efectividad del mismo. Debido a que de forma general en la literatura son escasos los artículos que tratan la selección y validación de modelos estadísticos, los formalismos que se requieren, o mostrar simplemente el paso a paso para la obtención y selección del modelo de regresión, planteamos el problema científico

Problema científico

Los estudios de análisis de enfriadoras de agua condensadas por aire carecen de metodologías para la confección y comparación de modelos de caja negra empleando el método de los mínimos cuadrados.

Hipótesis

La elaboración de una metodología que permita la confección de modelos matemáticos de caja negra que permita generar varios tipos de modelos y seleccionar la mejor variante

Objetivo general

Construir varios modelos matemáticos empleando la técnica de los mínimos cuadrados y comparar y seleccionar los mismos a través de métricas estadísticas. Comparar la precisión de modelos de en el funcionamiento de una enfriadora.

Objetivos específicos

- 1- Analizar el estado del arte de la construcción de modelos matemáticos tipo caja negra para enfriadoras condensadas por aire.
- 2- Elaborar la metodología que emplee técnicas estadísticas de regresión lineal en la construcción de modelo enfriadora.
- 3- Comparar y seleccionar el modelo que refleje la mejor bondad del ajuste y aplicando además el principio de parsimonia

Tareas

- 1- Seleccionar variables dependientes e independientes de la caja negra para enfriador condensado por aire.
- 2- Construir diferentes tipos de modelos (univariable, bivariable, orden 2
- 3- Utilizar el paquete Statgraphics, que refleje la mejor bondad del ajuste.
- 4- Analizar las variables de temperatura de entrada y salida para una enfriadora condensa por aire
- 5- Validar los modelos a partir de datos proporcionados por el fabricante.

CAPITULO I Estado del arte del modelado energético de enfriadoras de agua.

1.1 Enfriadoras de agua. Fundamentos básicos.

Una enfriadora de agua o chiller (terminología anglosajona) es un tipo de sistema de climatización que permite satisfacer la demanda térmica de acondicionamiento del espacio interior de las edificaciones. Estos se caracterizan fundamentalmente por emplear agua como fluido calo-portador, extrayendo el calor del medio y por consiguiente disminuyendo la temperatura del aire mediante un proceso de termo-transferencia. Ver figura 1

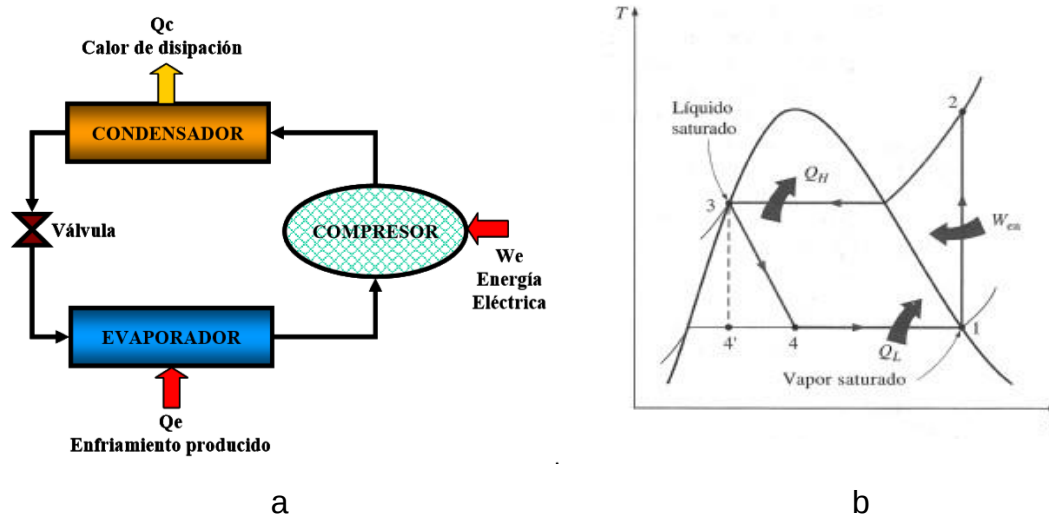


Fig. Principio de funcionamiento de los **principales** dispositivos del sistema de refrigeración a) componentes del sistema. b)

En el ciclo de compresión de vapor, el refrigerante es comprimido y después condensado a una elevada presión y temperatura, transfiriendo calor al ambiente o al agua de refrigeración en el condensador. Seguidamente el refrigerante condensado a alta presión es retornado a través de la válvula de expansión al evaporador, para absorber calor una vez más desde el agua a refrigerar. En ambos ciclos, el refrigerante líquido fluye dentro del evaporador y absorbe calor desde el agua a enfriar, y se evapora a una temperatura y presión baja.

Estos sistemas comparados con los sistemas de climatización individual, ofrecen muchas ventajas relacionadas con el funcionamiento, la confiabilidad y su eficiencia. La inversión inicial es alrededor de 2.2 veces mayor que los sistemas de climatización individual, pero el confort que se logra es superior

tanto por su operación más silenciosa, como desde la estética interior y exterior de la edificación. La distribución térmica requiere menos espacio en comparación que los sistemas Todo-Aire, siendo muy adecuado para edificios con espacio limitado. Permiten controlar simultáneamente las diferentes zonas térmicas, así como los circuitos de climatización y calefacción. Finalmente, estos sistemas, se emplean en grandes hoteles, restaurantes, cines, teatros, centros comerciales, hospitales y en otros edificios públicos. Suelen instalarse durante la construcción del edificio, aunque en algunos casos se instalan durante remodelaciones de edificios existentes

Las máquinas enfriadoras de agua o chiller (terminología anglosajona) son el eje central de una planta, que a su vez están compuesta por varias máquinas y son las responsables del 60 % del consumo energético del sistema de climatización(Chan, 2015). El chiller puede accionarse mediante el ciclo mecánico de compresión de vapor o un ciclo térmico de absorción. Las máquinas que emplean compresores pueden clasificarse de forma general en: compresores reciprocantes, centrífugos, de tornillo, de espiral. En el caso de los de ciclo de absorción, pueden ser de Absorción o Adsorción. Estos enfriadores están disponibles en diferentes rangos en términos de capacidad de enfriamiento según el tipo de compresor específico y el coeficiente de rendimiento (COP), como se muestra en la tabla 1.

Tabla 1.1: Rango de capacidad de frío y eficiencia de las enfriadoras de agua disponibles en el mercado(Navigant Consulting, 2014)

Tipo de Chiller	Rango de capacidad de frío (kW)	Eficiencia (COP)
Absorción	10 a 20 500	0.5-1.1
Adsorción	19 a 1 000	0.5-0.65
Centrífugos	280 a 14 000	6.1
Reciprocantes	7 a 1600	2.81
Tornillo	100 a 4 400	2.84
Espiral	528 a 1 056	3.02

Las enfriadoras se clasifican además por su forma de rechazar el calor hacia el ambiente. Las mismas pueden ser condensadas por agua o por aire. En términos de eficiencia los sistemas condensados por agua ofrecen ventajas sobre los sistemas condensados por aire, por ejemplo, según(Kaya & Alidrisi, 2016) los chiller condensados por aire consumen aproximadamente 1.40-1.75 kW/t, mientras que un chiller condensado por agua que posee el mismo rango

de capacidad puede disminuir su consumo en el rango de 0.46-0.65 kW/t. Además diferentes normas establecen que el rango de mínimo COP bajo condiciones de carga total para los chillers enfriados por aire deben variar entre 2.40-3.06 y los chillers condensados por agua entre 3.80-6.39 respectivamente (YU & Chan, 2010). Por otro lado, estos se pueden encontrar en un rango de capacidad más amplio entre 10 hasta 3 000 Toneladas de refrigeración (TR), y los condensados por aire entre 7 hasta 500 TR. Otra ventaja de estos sistemas es que poseen un tiempo de vida útil mayor de 20 a 30 años, mientras que los condensados por aire de 15 a 20 años.

Los condensados por agua requieren de una torre de enfriamiento además de un conjunto de subsistemas como son: alimentación de agua y drenaje, tratamiento de agua, filtros, sistemas de bombeo, y un sistema de control extra. Esto provoca que sean de mayor complejidad, además que se incrementan los costos de instalación y mantenimiento, favoreciendo en algunos casos la selección de los sistemas condensados por aire. Estos últimos se comercializan como sistemas compactos, optimizándose el tiempo para su diseño, instalación y montaje. Por otro lado, cuando en la zona el agua es escasa los sistemas condensado por agua no son recomendables, también su eficiencia disminuye cuando los niveles de humedad relativa y temperatura son elevados(Kumbar & N.Pandure, 20017)

Se destaca además que debido a que las torres de enfriamiento constituyen un riesgo de proliferación de la Legionelosis se emplean regulaciones para su estricto control y explotación(ASHRAE, 2017). Finalmente se recomienda que cuando no se tiene ni el presupuesto, ni las posibilidades óptimas para el mantenimiento, pues emplear chillers condensados por aire cuando estamos en presencia de plantas de 200 Ton de refrigeración de capacidad y menos. (Stanford III, 2016)

1.2. Evaluación energética del funcionamiento de un chiller

La modelación de un sistema de climatización centralizada implica elaborar modelos matemáticos de los distintos componentes (por ejemplo, la enfriadora, el sistema de bombeo, el sistema de condensación, etc.) y vincular el mismo con las dinámicas térmicas de la edificación y las condiciones del contorno donde el

sistema se desarrolla, por ejemplo, las condiciones climatológicas.(Underwood & YiK, 2004). Los resultados de este modelo serán el comportamiento y/o predicción de parámetros energéticos, eficiencia del sistema o posibles estados termodinámicos.

Para analizar un sistema en general, cada uno de sus sub-componentes puede considerarse un sistema en sí. Los modelos matemáticos de los mismos se derivande las leyes físicas que rigen los procesos que tienen lugar y del estado de las sustancias de trabajo dentro y fuera de los límites de su funcionamiento. Para determinar en tiempo real el consumo energético de un chiller se pueden emplear tres tipos de modelos matemáticos: modelos de caja blanca, modelos de caja negra y modelos de caja gris.

1.2.1. Modelos de caja blanca

Los modelos de caja blanca son aquellos donde se valora de forma individual cada uno de los componentes de un chiller a través de las ecuaciones que gobiernan los procesos físicos, termodinámicos y de transferencia de calor en el sistema.

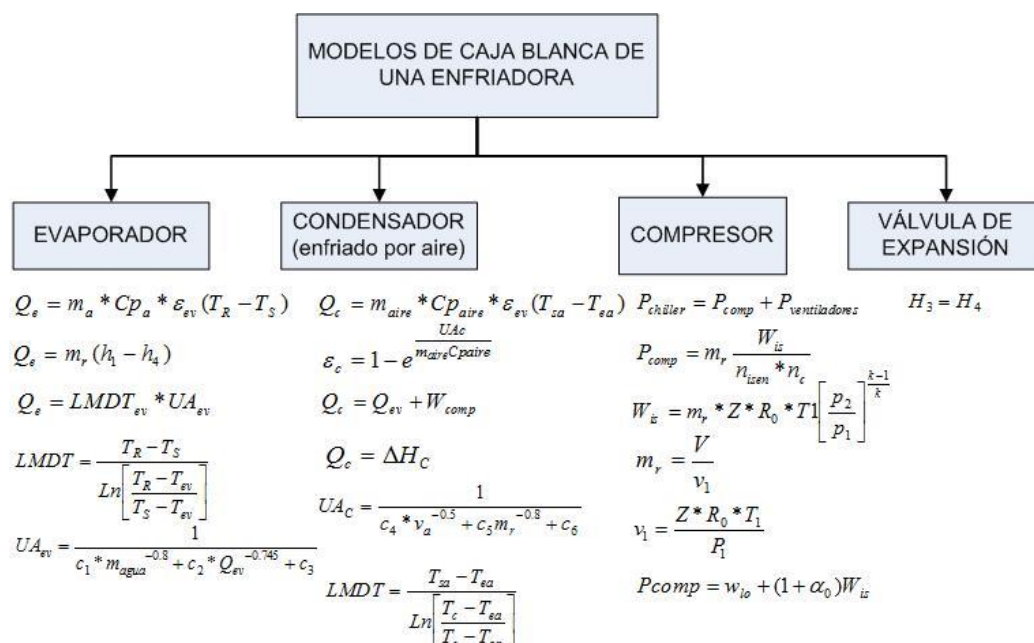


Figura: 1Modelo de caja Blanca(Ledesma, 2013)

Este tipo de modelo posee ventajas y limitaciones, por ejemplo, su predicción tiene un significado físico y los errores en el modelado son más fáciles de detectar, fundamentalmente los relacionados con las entradas (las condiciones

del contorno, con la solución numérica, iteraciones, etc.), permitiendo una depuración del modelo, logrando que éste se parezca más a las condiciones reales. Otra ventaja es que puede ser aplicable a componentes similares. Sus desventajas consisten en el tiempo necesario para su elaboración, lo complejidad, pues a menudo requieren de la entrada de un gran conjunto de parámetros para la descripción de las características físicas del sistema y sus componentes. Esta, a su vez, puede limitar sus aplicaciones, ya que la información requerida puede ser difícil para obtener. (Li & Wen, 2014)

1.2.2. Modelos de caja negra

Una forma alternativa de configurar los modelos matemáticos son los modelos de caja negra. Estos son modelos empíricos que solo dependen de parámetros de operación de un chiller. Se requieren datos del fabricante o mediciones efectuadas, ambos deben incluir un amplio rango de condiciones de operación de estas máquinas (Taylor, 2012). El modelo de caja negra relaciona numéricamente los resultados o salidas con las entradas más influyentes. La variable dependiente o de salida se establece en dependencia del objetivo del estudio. Por ejemplo, COP (Tingting, L, Xia, & Zhihui, 2019), (Chan, 2015); Potencia eléctrica (Ardakani et al, 2008), (Beghi et al, 2011), (Santos et al, 2014). Capacidad nominal, (Seo et al, 2016), Temperatura Set point (Romero & J.M. Belman-Flores, 2011), entre otros.

Así mismo las variables independientes o predictoras pueden ser diversas, se pueden emplear de forma individual o combinadas. Por ejemplo, (Wei & G. Kusiak, 2014) determinaron el consumo de energía eléctrica a través de las variables: flujo másico de agua helada, diferencia de temperatura de entrada y salida del condensador y la entalpía del aire. El modelo construido por (Yi K & Lam, 1998) empleó el coeficiente de carga parcial o razón de carga (PLR) y la temperatura del agua a la entrada del condensador. (Cheng & Ch. Wang, 2015) le agregaron además la temperatura del set point. (Chang, C.Y. Lu, J.T. Lee, & T.S. Chen, 2013) sustituyeron el PLR por la capacidad de frío. Otras variables predictoras que se han empleado en la construcción del modelo han sido: La velocidad del compresor (Romero & J.M. Belman-Flores, 2011) variables climatológicas (temperatura exterior y humedad relativa) (Le Cam & R. Zmeureanu, 2013); (Rossi & J.E. Braun., 1997) además se ha empleado

información más específica como la hora del día, (Le Cam et al. 2013). Ambas variables, dependientes e independientes se derivan en una curva, que emplean una serie de parámetros o coeficientes de correlación que permiten su ajuste.

Por la naturaleza de este tipo de modelo, son más convenientes en la ingeniería. No obstante, sus restricciones están dadas en que no permite interpolar datos que se encuentren fuera del rango (Wang, H. 2017) por lo que no puede reflejar los efectos de cualesquiera factores cuya influencia fue omitida en su derivación. Sus resultados además no explican en detalle el fenómeno físico en que la tecnología incurre. Los modelos de caja negra se pueden construir de diversas formas. Empleando la regresión lineal múltiple (Jayasekara, S. y autores 2014) (Seo and Lee, 2016), (Wang, H. 2017) (Ardakani y autores, 2008); (Chang, Yung-Chung, 2004); (Chang, Yung-Chung 2007); (Santos y autores, 2014); (Beghi A. y autores, 2011); (Seo y Lee, 2016); Modelo de regresión de Cox (Yu, F.W. y Ho, H.T. 2019); Regresión de Kernel (Le Cam, M y autores 2013); Redes neuronales (Swider y autores, 2001); (Kusiak, A. y Xu, G. 2012); (Jia Z. y Zhao T. 2018) Modelos neuro-difusos (Attia y autores, 2015); Modelos auto-regresivos de movimiento (Romero y autores, 2011); Modelos de caja Jenkins (Romero y autores, 2011); Algoritmos genéticos (Wang, L. y autores, 2018); Least squares support vector machine (Han, H. y autores, 2019) (Namburu y autores, 2005).

1.2.3 Modelos de caja gris

Los modelos de caja gris es una mezcla entre el modelo de caja blanca y negra. En este caso también se basa en los principios termodinámicos, los parámetros de funcionamiento de un chiller y los coeficientes de los modelos tienen un significado físico claro. Este tipo de modelo es más preciso. Se pueden emplear para generar parámetros de funcionamiento a través de la extrapolación, pero requieren de numerosos parámetros que son difíciles de obtener en caso de una planta de chiller en operación real o de datos del fabricante. Se generan a través de plantas pilotos. Estos modelos son más factibles para análisis de diseño, control, simulación y detección de fallas (Wang, Handong, 2017). Los modelos de caja gris más representativos son el modelo universal de Gordon-Ng (Ng y autores, 1996) y el modelo simplificado

de Lee (Lee, T.S, 2004). Otros estudios efectuados con este modelo se encuentran en (Browne, M.W. y Bansal, P.K., 1998) (Koury, R.N.N. y autores, 2001) (McIntosh, I.B.D. y autores, 2000) (Browne, M.W. y Bansal, P.K., 2002) (Cecchiato, L. y autores, 2010) (Yao, Y. y autores, 2015)

$$\frac{T_{ee}}{T_{ec}} \left(1 + \frac{1}{COP} \right) - 1 = \frac{T_{ee}}{Q_e} \Delta S_T + Q_p \frac{(T_{ec} - T_{ee})}{T_{ec} * Q_e} + \frac{R * Q_e}{T_{ec}} \left[1 + \frac{1}{COP} \right]$$

Donde:

$T_{ee} \rightarrow$ Es la temperatura de entrada del agua al evaporador; K

$T_{ec} \rightarrow$ Es la temperatura de entrada del agua al condensador; K

$COP \rightarrow$ Coeficiente de trabajo del ciclo.

$Q_e \rightarrow$ Calor absorbido en el evaporador; kW

$Q_p \rightarrow$ Pérdidas de calor equivalente; kW

$\Delta S_T \rightarrow$ Generación de entropía total del sistema; $\frac{kW}{K}$

$R \rightarrow$ Resistencia térmica total del intercambiador; $\frac{W}{m^2 * K}$

$$R = \frac{1}{\varepsilon_c * m_c * c_p} + \frac{1}{\varepsilon_e * m_e * c_p}$$

$\varepsilon_c \rightarrow$ Efectividad del condensador

$\varepsilon_e \rightarrow$ Efectividad del evaporador

$m_c \rightarrow$ Flujo de agua de enfriamiento; $\frac{kg}{s}$

$c_p \rightarrow$ Calor específico; $\frac{kJ}{Kg^0C}$

1.3 Modelos de caja negra basados en la regresión lineal en problemas

Para la solución de problemas de optimizaciones usualmente se emplea el modelo de regresión lineal. Algunos autores se basan en la curva COP-PLR o COP-Pot para establecer sus modelos, porejemplo: Chang, Yung-Chung (2004), Chang et al (2005), Ardakani et al, 2008; Santos et al, 2014), (Cheng, Q. y autores, 2017), (Liao, Y. y autores, 2018). Porotrolado (Huang et al 2016) refiere que establecer el control basado en el PLR-COP es complejo cuando se trata de aplicaciones reales debido a que esta variable solo puede ser controlada indirectamente, por lo que emplean en sus modelos variables temperatura a la salida del evaporador, la temperatura del agua a la entrada y salida del condensador para el caso de las enfriadoras condensadas por agua y la temperatura del aire a la entrada del condensador para las enfriadoras condensadas por aire. (Seo and Lee, 2016) (Wang, Y. y autores, 2018). Otros también emplean el flujo másico del agua helada para el caso de chiller de flujo variable (Beghi A. et al, 2011). Ejemplos de modelos caja negra de problema de optimización se encuentran en la tabla:

Tabla1 : Modelos matemáticos de caja negra empleados en la simulación de enfriadoras de agua

Ardakani et al, 2008	$P_i = a_i + b_i * PLR + c_i * PLR^2 + d_i * PLR^3$
Chang, Yung-Chung (2004)	$COP_i = a_i + b_i PLR + c_i PLR^2$
Chang et al (2005)	$COP_i = a_i + b_i (PLR) + c_i (PLR)^2$
Chang, Yung-Chung(2007)	$P_{ch,i} = k_{0,i} + k_{1,i} \Delta T_{cw} + k_{2,i} \Delta T_{cw}^2 + k_{3,i} \Delta T_{chw} + k_{4,i} \Delta T_{chw}^2 + k_{5,i} \Delta T_{cw} \Delta T_{chw}$
Santos et al, 2014	$P_i = a_i + b_i PLR + c_i (PLR)^2 + d_i (PLR)^3$
Beghi A. et al 2011	$Pe_{full\ i} = a_{ei} + b_{ei} T_{chwr} + c_{ei} T_{air} + d_{ei} \dot{m}_w$ $Pc_{full\ i} = a_{ci} + b_{ci} T_{chwr} + c_{ci} T_{air} + d_{ci} \dot{m}_w$
Seo and Lee, 2016	$ChillerCapFtemp = a + b(T_{cw,l}) + c(T_{cw,l})^2 + d(T_{cond,e}) + e(T_{cond,e})^2 + f(T_{cw,l})(T_{cond,e})$ $ChillerEIRFtemp = a + b(T_{cw,l}) + c(T_{cw,l})^2 + d(T_{cond,e}) + e(T_{cond,e})^2 + f(T_{cw,l})(T_{cond,e})$

El nivel de complejidad, así como el total de variables predictores empleados en la construcción del modelo de regresión lineal, dependerá de los datos, así como, de los resultados de las pruebas de calidad al modelo. La diversidad en las entradas y salidas del modelo, la complejidad del blackbox chiller era muy

diferente. Por ejemplo, Yik et al. (1998) el modelo tenía dos variables predictoras y nueve parámetros, Reddy et al. (2002) el modelo tenía dos predictores y 10 parámetros, el modelo de Cui y Wang (2005) tenía tres predictores y ocho parámetros de regresión, y Hydeman et al. (2002) el modelo tenía 5 predictores y 19 parámetros.

En el modelado basado en datos, es esencial el equilibrio de la luz entre el número de predictores y los parámetros para evitar el ajuste o el ajuste de los ajustes (Gunay et al, 2016). En muchos casos, la evaluación de modelos de modelos de frío se realizó en una parte de los datos que no se utilizaron en el entrenamiento modelo (p. Ej., Validación cruzada; Rossi y Braun 1997; Haves y Khalsa 2000; Swider y otros. 2001; Reddy and Andenens2004; Hydeman y Gillespie2002; Hydeman y otros. Li y Braun 2003; Swider2003; Namburu et al. 2005; Romeo y Gareta2006; Namburu et al. 2007; Reddy2008; O'Neill et al. 2011; Romero et al. 2011; Labus et al. 2013; Le Cam et al. 2013; Zhao et al. 2013; Zhao et al. 2013; Beghi et al. 2016). Sin embargo, en algunos casos, los indicadores estadísticos (p. Ej., El coeficiente de determinación $[R^2]$, RMSE, el coeficiente de variación de los datos) se calcularon utilizando el conjunto de entrenamiento en sí mismo, principalmente debido a las limitaciones del tamaño del conjunto de datos (Haves y Khalsa2000; Labus et al. 2013; Yik y Lam, 1999). La Norma 14 de ASHRAE (2014) presenta las mejores prácticas para desarrollar modelos de regresión (para estimar los acontecimientos más importantes en la construcción de proyectos de gestión de energía); sin embargo, muchos temas importantes que se relacionan con la selección de modelos estadísticos y la validación deben estar cubiertos (por ejemplo, modos de modelado de modos de caja negra y gris, validación cruzada de sub-muestreo aleatorio, y regresión hacia adelante / hacia atrás).

Conclusiones parciales

1. Se ha podido observar en la revisión bibliográfica que las modelaciones de los sistemas de producción de frío se dirigen hacia tres tendencias principales:

la obtención de los modelos a partir de ecuaciones termodinámicas y de transferencia de calor (determinísticos), a partir de análisis de regresión fundamentados en el comportamiento de los flujos de entrada y salida de cada componente o del sistema (caja negra)

2.Los sistemas de climatización centralizada son utilizados en grandes instalaciones, pues resulta más factible centralizar la distribución de frío a los recintos a climatizar que ubicar unidades independientes en cada uno de ellos.

CAPITULO II Materiales y métodos

En la Fig. 2.1 se describe gráficamente la metodología general orientada a la evaluación de los modelos matemáticos empleando la selección del mejor modelo de regresión, los aspectos conceptuales y los procedimientos de cálculo que se utilizan en el presente trabajo están basado principalmente en la literatura de Elemento de Econometría.

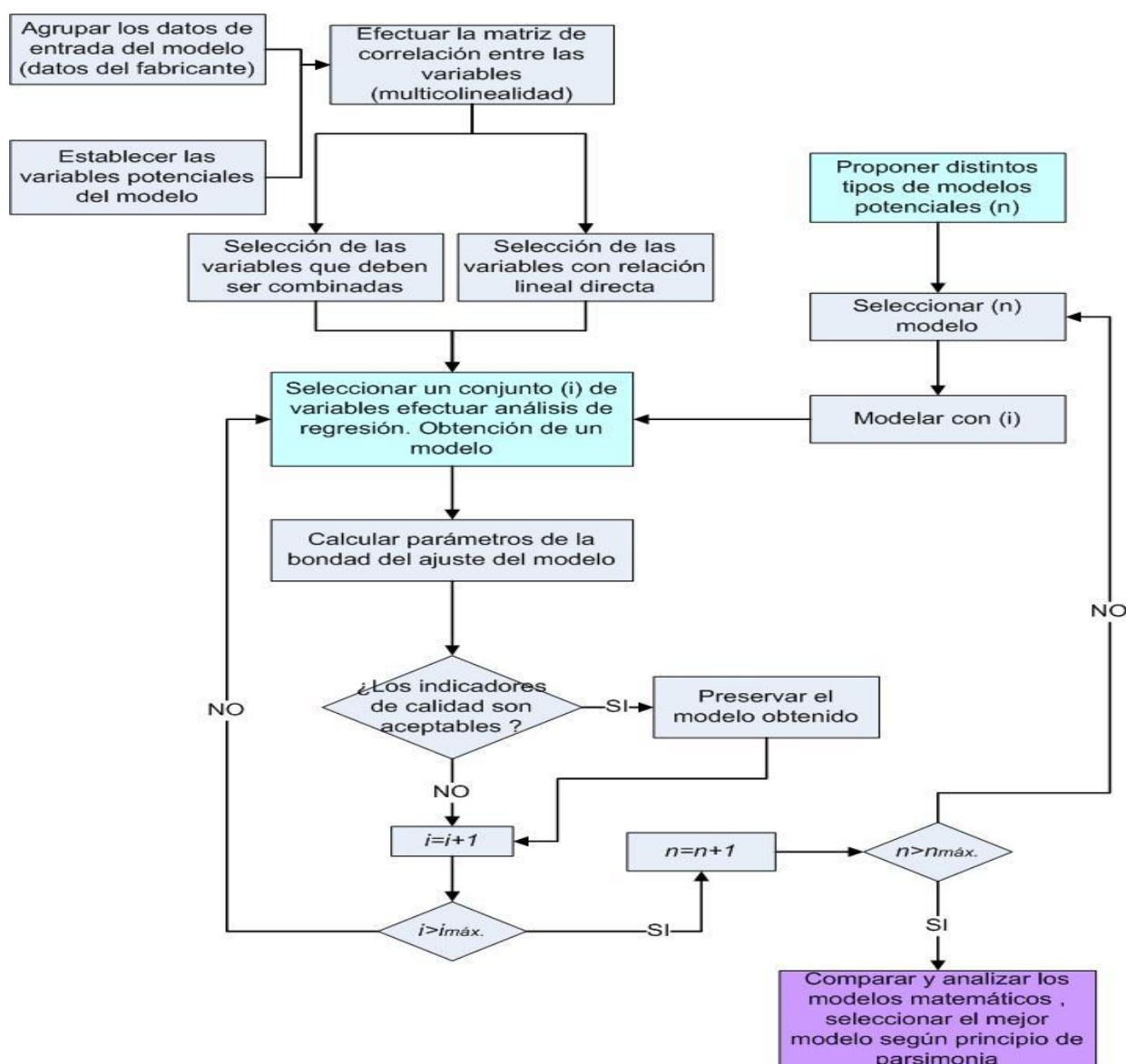


Fig.2.1 Esquema metodológico para elaboración de modelos de caja negra.
Elaboración propia

2.1 Fases de construcción de un modelo

En muchos casos la construcción o creación de modelos matemáticos útiles sigue una serie de fases bien determinadas (ARTEAGA J. A., 2010)

1. **Identificación** de un problema o situación compleja que necesita ser simulada, optimizada o controlada y por tanto requeriría un modelo matemático predictivo para hacer efectivo el mismo
2. **Elección del tipo** de modelo, esto requiere precisar qué tipo de respuesta pretende obtenerse, cuáles son los datos de entrada o factores relevantes, y para qué pretende usarse el modelo. Esta elección debe ser suficientemente simple como para permitir un tratamiento matemático asequible con los recursos disponibles.
3. **Formalización** del modelo en la que se detallarán qué forma tienen los datos de entrada, qué tipo de herramienta matemática se usará, como se adaptan a la información previa existente. También podría incluir la confección de algoritmos, ensamblaje de archivos informáticos, etc. En esta fase posiblemente se introduzcan también simplificaciones suficientes para que el problema matemático de modelización sea tratable computacionalmente.
4. **Comparación de resultados:** los resultados obtenidos como predicciones necesitan ser comparados con los hechos observados para ver si el modelo está prediciendo bien. Si los resultados no se ajustan bien, es común volver a la fase 1.

Los modelos de regresión lineal son ampliamente usados en la ingeniería ya que sirven para analizar el comportamiento de las variables de entrada (o regresora) y salida (o respuesta) estableciendo predicciones y estimaciones.

2.1.1 Selección del modelo de regresión

2.1.1.1 Modelo de regresión lineal simple

Los modelos de regresión lineal son ampliamente usados en la ingeniería ya que sirven para analizar el comportamiento de las variables de entrada (o regresora) y salida (o respuesta) estableciendo predicciones y estimaciones.

En este trabajo la variable regresora corresponde a la distorsión armónica individual de corriente y la variable de respuesta corresponde a la distorsión armónica individual de tensión. La ecuación (2.1), muestra la representación de un modelo de regresión lineal simple, donde Y es la respuesta, X es la variable regresora, β_0 y β_1 son los parámetros del modelo o coeficientes de regresión y ε es el error del modelo. (Gómez, 2013)

$$Y = \beta_0 + \beta_1 * X + \varepsilon \quad (2.1)$$

La ecuación (2.2) muestra el modelo ajustado de regresión lineal simple, donde $\hat{Y}$ es el valor estimado o esperado de la variable de respuesta para la variable regresora X. Los estimadores de los parámetros del modelo son calculados por el método de los mínimos cuadrados usando las expresiones (2.3) y (2.4).

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 * x \quad (2.2)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 * \bar{x} \quad (2.4)$$

2.1.1.2 Modelo de regresión polinomial

Los modelos de regresión polinomial se usan cuando la variable de respuesta muestra un comportamiento curvilíneo o no lineal. La ecuación (2.5), describe el modelo de regresión polinomial de orden k en una variable regresora y la ecuación (2.6), muestra el modelo ajustado de regresión polinomial de orden k. Los estimadores de los parámetros del modelo se obtienen por el método de los mínimos cuadrados usando la ecuación (2.7), donde $\mathbf{y}$, $\mathbf{X}$, $\mathbf{X}^T$ son vectores. En este trabajo se aplican los modelos de regresión polinomial de orden 2 y orden 3 en una variable regresora.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k + \varepsilon \quad (2.5)$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + \dots + \hat{\beta}_k X^k \quad (2.6)$$

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.7)$$

2.1.1.3 Modelo de regresión lineal múltiple

La ecuación (2.8), describe un modelo de regresión lineal múltiple y la ecuación (2.9) muestra el modelo ajustado de regresión lineal múltiple. En este modelo la variable de respuesta se puede relacionar con k variables regresoras. Se justifica aplicar el modelo de regresión lineal múltiple cuando interviene más de una variable regresora. Los estimadores de los coeficientes de este modelo se obtienen a partir de la ecuación (2.9). En este trabajo se consideraron 3 variables regresoras.

$$Y = \beta_0 + \beta_1 X + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon \quad (2.8)$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_K X_K \quad (2.9)$$

En los modelos de regresión lineal múltiple se debe evaluar la presencia de la multicolinealidad, que aparece cuando existe una dependencia casi lineal entre las variables regresoras del modelo. La multicolinealidad provoca problemas en la determinación de los estimadores de los coeficientes del modelo. La técnica más común que se usa para detectar la presencia de multicolinealidad es el análisis de la matriz de correlación.

Coeficiente de determinación R^2

El coeficiente de determinación R^2 mide la proporción de la variación de la respuesta Y que es explicada por el modelo de regresión. El coeficiente R^2 se calcula usando la ecuación (2.10), donde SSR es la medida de variabilidad del modelo de regresión y SST corresponde a la medida de variabilidad de Y sin considerar el efecto de la variable regresora X. (Astorga Gómez, 2014)

$$R^2 = \frac{SS_R}{SS_T} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}; 0 \leq R^2 \leq 1 \quad (2.10)$$

Si el valor de R^2 se acerca a 1, entonces se tiene un buen ajuste y el modelo de regresión es capaz de explicar adecuadamente la variación de la variable de respuesta Y.

2.2 Obtener el modelo

2.2.1 Calcular los coeficientes

Mediante un modelo de regresión lineal múltiple (MRLM) tratamos de explicar el comportamiento de una determinada variable que denominaremos variable a explicar, variable endógena o variable dependiente, (y representaremos con la letra Y) en función de un conjunto de k variables explicativas $X_1, X_2, \dots, X_k$ mediante una relación de dependencia lineal (suponiendo $X_1 = 1$):

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.11)$$

Siendo ε el término de perturbación o error.

Para determinar el modelo anterior, es necesario hallar (estimar) el valor de los coeficientes $\beta_1, \beta_2, \dots, \beta_k$. La linealidad en parámetros posibilita la interpretación correcta de los parámetros del modelo. Los parámetros miden la intensidad media de los efectos de las variables explicativas sobre la variable a explicar y se obtienen al tomar las derivadas parciales de la variable a explicar respecto a cada una de las variables explicativas:

$$B_j = \frac{\partial Y}{\partial X_j}; j = 1, \dots, k \quad (2.12)$$

Nuestro objetivo es asignar valores numéricos a los parámetros $\beta_1, \beta_2, \dots, \beta_k$. Es decir, trataremos de estimar el modelo de manera que, los valores ajustados de la variable endógena resulten tan próximos a los valores realmente observados como sea posible

Método de los mínimos cuadrados ordinarios (MCO): estimadores

La suma de los residuos -unos positivos, otros negativos y algunos incluso nulos- es de esperar que sea cero; por ello como medida de la pequeñez total de los errores se utiliza la suma de sus cuadrados, que son siempre positivos. Precisamente, uno de los métodos de estimación más usados en el análisis de regresión se basa en obtener como estimaciones de los parámetros del modelo, β_j , aquellos valores b_j que hagan mínima la suma de los cuadrados de los residuos; por ello éste es denominado Método de los Mínimos Cuadrados Ordinarios (MCO). Para desarrollar el método de los mínimos cuadrados se

debe, por supuesto, derivar la suma de cuadrados de los residuos, $\sum (y - \bar{y})^2$ respecto a los distintos b_j , igualar a cero estas derivaciones -como condición de mínimo-, y despejar del sistema de ecuaciones resultante los valores de las b_j .

Al aplicar el método MCO al modelo de regresión simple $Y = \beta_0 + \beta_1 X + \varepsilon$ se obtienen los estimadores:

$$b_1 = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \text{ y } b_0 = \frac{\sum y - b_1 \sum x}{n} \quad (2.13)$$

2.2.2 Análisis de varianza o prueba F en la regresión

La prueba más utilizada en el análisis de modelos de regresión es la llamada prueba F o análisis de varianza en la regresión. La misma pretende verificar, con determinado nivel de significación previamente escogido, si el modelo desarrollado alcanza un grado suficiente de precisión estadística, o lo que es lo mismo, si tiene significatividad o utilidad predictiva.

La formulación de la prueba F en una regresión simple se basa en el hecho de que si el coeficiente asociado a la variable explicatoria en el modelo, β_1 , no es significativamente distinto de cero, entonces la variable dependiente, Y , no va a depender significativamente de dicha variable explicatoria, X , y por tanto el modelo no es significativo. En lenguaje estadístico queda:

$H_0 : \beta_1 = 0$ (modelo no preciso, no significativo...)

$H_1 : \beta_1 \neq 0$ (modelo significativo)

Como estadístico de prueba se utiliza el cociente entre dos estimadores distintos para la varianza asociada al término de perturbación. Estos estimadores reciben, entre otros, los nombres de varianza explicada y varianza residual. El cálculo de dicho estadístico, similar al del análisis de varianza para comparar medias de diferentes poblaciones, se suele presentar en forma de una tabla, llamada tabla de análisis de varianza para la regresión (tabla ADEVA, o ANOVA del inglés *analysis of variance*); la estructura de la tabla es la siguiente:

Tabla 2.1 ANOVA				
Fuente de Variación	Sumas de Cuadrados	Grados de Libertad	Varianzas o Cuadrados Medios	Estadígrafo
explicada	SCE	p - 1	$S_E^2 \equiv CME = \frac{SCE}{p-1}$	$F_0 = \frac{S_E^2}{S_R^2} = \frac{\frac{SCE}{p-1}}{\frac{SCR}{n-p}}$
residual	SCR	n - p	$S_R^2 \equiv CMR = \frac{SCR}{n-p}$	
total	SCT	n - 1		

Es bueno señalar que las hipótesis planteadas, en total analogía con el análisis de varianza tradicional, son equivalentes a decir:

$$H_0 : \frac{E(S_E^2)}{E(S_R^2)} = 1 \quad H_0 : \frac{E(S_E^2)}{E(S_R^2)} > 1 \quad (2.14)$$

Ello se debe a que S_E^2 es en general un estimador sesgado de la varianza total, y sólo es insesgado si la hipótesis nula es cierta, mientras que S_R^2 es siempre un estimador insesgado. El estadígrafo calculado es conocido como estadígrafo o estadístico F ya que por tratarse de un cociente de varianzas sigue una distribución F de Fisher. Los grados de libertad, como se ve en la tabla, son p - 1 para el numerador y n - p para el denominador. Y tratándose de una regresión simple siempre será p=2. Por tanto, la región crítica de la prueba tiene la forma:

$$W = \{F_0 : F_0 > F_{1-\alpha}(p-1; n-p)\} \quad (2.15)$$

Alternativamente, si se calcula el p-valor, probabilidad asociada o probabilidad marginal del estadígrafo F lo cual en general se hace sólo si está trabajando con un software- entonces la región crítica puede expresarse como:

$$W = \{p(F_0) : p(F_0) > \alpha\} \quad (2.16)$$

Como conclusión, si se rechaza la hipótesis nula puede asumirse que el modelo es estadísticamente significativo.

2.2.3 Prueba t en la regresión

Otra prueba de uso casi absoluto en el análisis de regresión es la llamada “prueba t”, por la forma del estadístico de prueba.

En el caso de la regresión, asumiendo el cumplimiento del supuesto de normalidad en las perturbaciones, puede probarse que los estimadores b_j , correspondientes a los distintos parámetros β_j , están asociados en su comportamiento probabilístico a una distribución t de Student con $n - p$ grados de libertad, siendo -evidentemente- $p=2$ para la regresión simple. Estrictamente hablando, quien sigue una distribución t de Student es la transformación llamada estudentización de los estimadores:

$$\frac{b_j - \beta_j}{S_{B_j}} \sim t(n - p) \quad (2.17)$$

Conocer esto permite de manera inmediata usar tal transformación, conocida en este contexto como estadígrafo t, para verificar a partir de la estimación obtenida si el verdadero valor de un parámetro β_j dado se puede asumir como efectivamente mayor, menor o igual que un cierto valor B_j considerado de interés o histórico. (Aguilera, 2010)

Las distintas alternativas de pruebas con su correspondiente región crítica quedan

Hipótesis que se plantea Región crítica

$$\begin{array}{ll} H_0: \beta_j = B_j & \\ H_1: \beta_j > B_j & W = \{t_{b_j}: t_{b_j} > t_{1-\alpha(n-p)}\} \end{array} \quad (2.18)$$

$$\begin{array}{ll} H_0: \beta_j = B_j & \\ H_1: \beta_j < B_j & W = \{t_{b_j}: t_{b_j} < -t_{1-\alpha(n-p)}\} \end{array} \quad (2.19)$$

$$\begin{array}{ll} H_0: \beta_j = B_j & \\ H_1: \beta_j \neq B_j & W = \{t_{b_j}: t_{b_j} < -t_{1-\alpha/2(n-p)}\} \text{ ó } \\ & \left\{ t_{b_j} > t_{1-\alpha/2(n-p)} \right\} \text{ ó } \\ & \{t_{b_j}: |t_{b_j}| > t_{1-\alpha/2(n-p)}\} \end{array} \quad (2.20)$$

2.2.4 Análisis de los supuestos de la regresión

Como ya se ha mencionado, la estimación de los modelos de regresión lineal mediante el método de los mínimos cuadrados se basa en una serie de supuestos simplificadores de la realidad. Una forma bastante extendida de hacer referencia a tales supuestos es la siguiente:

- media condicional nula en las perturbaciones
- no multicolinealidad
- homocedasticidad en las perturbaciones
- no autocorrelación entre las perturbaciones
- normalidad en las perturbaciones

A continuación, se explican detalladamente cada una de ellas:

2.2.4.1 Supuesto de media condicional nula en las perturbaciones

El supuesto de media condicional nula en las perturbaciones establece que la media o valor esperado de las perturbaciones es cero para todos los posibles valores de las variables explicatorias. Matemáticamente se expresa como:

$$E(u_i / X_{1i}, X_{2i}, \dots, X_{ki}) = 0 \quad (2.21)$$

En aras de evitar confusiones durante la confrontación bibliográfica conviene saber que el planteamiento de este supuesto se encuentra frecuentemente subdividido en dos: supuesto de media nula –no condicionada– en las perturbaciones, y supuesto de exogeneidad de las variables explicatorias. En estos términos lo que se exige es lo siguiente:

- media nula de u : $E(u) = 0$

(La media o valor esperado global de las perturbaciones es cero.)

- exogeneidad de las X_j : $\text{Cov}(X_{ji}, u_i) = 0$

(Las variables explicatorias no están correlacionadas con el término de perturbación)

2.2.4.2 Supuesto de no multicolinealidad

Se le llama multicolinealidad a la presencia alta correlación entre las variables explicatorias de un modelo, y el caso extremo en que la relación entre ellas es exacta es llamado multicolinealidad perfecta.

La existencia de multicolinealidad perfecta en un modelo impide la estimación de los parámetros del mismo, pues en la matriz de las X habría columnas

linealmente dependientes, lo cual anula el determinante e imposibilita el cálculo de la inversa de $X^T X$.

En el caso de multicolinealidad no perfecta pero alta, aunque los estimadores mantienen las propiedades de insesgadez y eficiencia, sus errores estándar, varianzas y covarianzas son más grandes y como consecuencia los intervalos de confianza tienden a ser más amplios y los estadísticos t perder significación.

Por otra parte, el hecho de que haya variables explicatorias dependientes entre sí hace que se pierda claridad en la interpretación de los parámetros del modelo. En efecto, b_k indica la influencia sobre Y de un aumento unitario de X_k cuando todas las demás variables explicatorias se mantienen fijas, pero si X_j está correlacionada con X_k es imposible aumentar X_k manteniendo fijo X_j .

En el análisis de un modelo la presencia de un R^2 alto, o lo que es equivalente, de una prueba F significativa, acompañado esto de pruebas t no significativas es un indicio de posible presencia de multicolinealidad; como se suele decir, de sospecha de multicolinealidad. También provoca sospecha de multicolinealidad la falta de significación teórico-práctica de algún estimador, en particular que éste resulte con signo distinto del esperado.

No obstante, la sospecha de multicolinealidad es sólo eso, sospecha; la certeza requiere del desarrollo de regresiones auxiliares entre las variables explicatorias y la determinación de si estas regresiones auxiliares son significativas, lo cual confirmaría una fuerte relación entre las variables involucradas, es decir, presencia de multicolinealidad en el modelo original. Para el caso particular de un modelo con dos variables explicatorias basta, sin embargo, con analizar el coeficiente de correlación entre dichas variables, r_{xy} , y si este es modularmente alto, por ejemplo $|r_{xy}| > 0.85$, la correlación entre las variables se considera suficientemente alta como para decir que hay elevada multicolinealidad.

2.2.4.3 Supuesto de homocedasticidad en las perturbaciones

Por homocedasticidad en las perturbaciones se entiende que las mismas mantengan una variabilidad constante a lo largo de todo el rango observacional del modelo. El incumplimiento de esto como supuesto de la regresión acarrea

severas consecuencias sobre los resultados del modelo, y por esta razón se han dedicado muchos métodos al análisis de la homocedasticidad.

En general se dice que existe homocedasticidad en un modelo de regresión cuando la varianza de los residuos es constante y que hay heterocedasticidad cuando no lo es.

Según el modelo clásico de regresión lineal, las perturbaciones aleatorias deben tener varianza constante a lo largo de todas las observaciones, matemáticamente esto se expresa como:

$$\forall i \ V(u_i) = \sigma^2 \quad \text{ó} \quad \forall i \ E(u_i^2) = \sigma^2$$

Gráficamente la homocedasticidad puede ilustrarse en la figura como sigue:

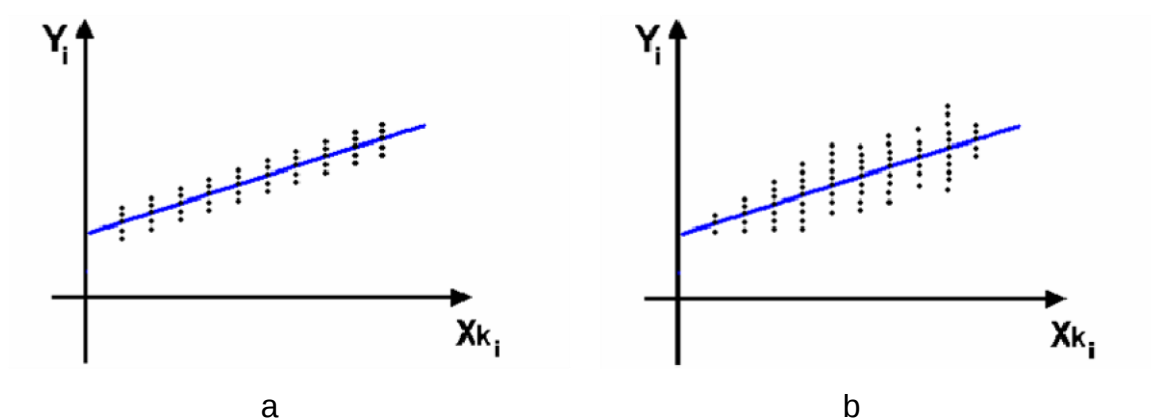


Figura : Representación gráfica de la variabilidad de los residuos a) Homocedasticidad, b) heterocedasticidad

La más inmediata y frecuente razón para la existencia de heterocedasticidad es que las variaciones en las variables explicatorias afectan no sólo el nivel sino también la dispersión –variabilidad- de la variable dependiente. Por ejemplo, a medida que aumentan los ingresos aumenta la oportunidad de elegir, y es común que aumente la variabilidad en el consumo; también, a medida que se estudia disminuyen los errores, y debe disminuir entonces la variabilidad en ellos.

Aunque se han desarrollado muchas pruebas estadísticas para verificar la presencia de homocedasticidad en los modelos de regresión, no

puededecirse que exista una definitiva o siempre útil. Entre las más empleadas cabe mencionar:

- Prueba de White
- Prueba de Park
- Prueba de Breusch-Pagan
- Prueba de Goldfeld-Quandt
- Prueba de Glejser
- Prueba de Bartlett

Prueba de White

Por la parte que le toca, la prueba de White es una de las pruebas más robustas usadas, pues no se basa en ningún supuesto sobre la naturaleza de la heterocedasticidad.

La hipótesis nula en la prueba de White, como es común en este tipo de prueba, es la existencia de homocedasticidad. El desarrollo de la misma requiere que se tengan los residuos del modelo estimado (e_i), como medida de la varianza de la perturbación correspondiente a cada observación; y que se efectúe una regresión auxiliar de dichos residuos al cuadrado frente a todas las variables explicatorias, los cuadrados de las mismas y, opcionalmente, los productos cruzados de ellas. Si en realidad hay homocedasticidad esta regresión auxiliar debe ser no significativa.

Para tomar una decisión utiliza como estadígrafo de prueba el producto de la cantidad de observaciones (n) por la R^2 de la regresión auxiliar de White, que bajo la hipótesis nula de homocedasticidad se distribuye con una distribución Chi-cuadrado cuyos grados de libertad son la cantidad de parámetros implicados en la regresión auxiliar de White sin contar el término de intercepto, o lo que es igual, la cantidad de términos con variables explicatorias en dicha regresión auxiliar (k_{Wh})

$$W = \{nR_{Wh}^2: nR_{Wh}^2 > \chi_{1-\alpha(k_{Wh})}^2\} \quad (2.24)$$

Así, por ejemplo, para un modelo con dos variables explicatorias la regresión auxiliar de White queda:

$$e_i^2 = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2 + v_i \quad (2.25)$$

Y las hipótesis de la prueba pueden plantearse como:

$$H_0: a_1 = a_2 = a_3 = a_4 = a_5 = 0 \text{ (hay homocedasticidad)}$$

$$H_1: \text{Al menos un } a_i \neq 0 \text{ (} i > 0 \text{) (hay heterocedasticidad)}$$

$$\text{Siendo la región crítica: } W = \{nR_{Wh}^2: nR_{Wh}^2 > \chi_{1-\alpha(5)}^2\}$$

Finalmente, como estas pruebas suelen desarrollarse a través de algún software estadístico o econométrico, y estos normalmente brindan como dato el p-valor asociado al estadígrafo de prueba, conviene tener presente que la región crítica expresada en función del mismo queda:

$$W = \{p(nR_{Wh}^2): p < \alpha\}$$

De existir heterocedasticidad en el modelo, dadas las consecuencias conocidas de ésta, se tendrá que proceder entonces a resolver esta situación

2.2.4.5 Supuesto de no autocorrelación entre las perturbaciones

Por autocorrelación entre las perturbaciones se entiende que las mismas guarden relación unas con otras a lo largo del rango observacional del modelo. La existencia de esta constituye un incumplimiento de los supuestos de la regresión, y acarrea severas consecuencias sobre los resultados del modelo, razón por la cual se han dedicado muchos métodos al análisis de la no autocorrelación. Según el modelo de regresión lineal, las perturbaciones aleatorias no deben estar correlacionadas, es decir:

$$\forall_{i \neq j} \text{cov}(u_i, u_j) = 0 \text{ ó } \forall_{i \neq j} E(u_i u_j) = 0 \quad (2.27)$$

Algunas de los factores que más frecuentemente conllevan la presencia de autocorrelación en un modelo son:

- la exclusión del modelo de variables que son necesarias
- la inadecuada especificación de la forma funcional del modelo

- la inclusión de retardos de las variables explicatorias en el modelo (modelos de retardos distribuidos)
- la inclusión de la variable dependiente en forma retardada como una variable explicatoria (modelos autorregresivos)
- la existencia de ciclos y tendencia en las series
- los errores de medición
- la manipulación previa de los datos (hallando medias móviles, por ejemplo)
- la proximidad espacial, en algunas series de datos de corte transversal

Se puede recurrir a diferentes pruebas estadísticas para analizar la existencia de autocorrelación en una serie; entre ellas:

- Prueba de Durbin-Watson
- Prueba de Breusch-Godfrey
- Prueba Ljung-Box o del estadístico Q
- Prueba χ^2 de independencia de los residuos
- Prueba de aleatoriedad, de corridas o de Geary

Prueba Durbin-Watson

Una de las pruebas más conocidas y populares en este contexto es la propuesta por Durbin y Watson, que parte del supuesto de que la autocorrelación debe a que los errores u_t son generados mediante un proceso o modelo autorregresivo de primer orden (lo que se llama un AR 1) o sea:

$$u_t = \rho u_{t-1} + \epsilon_t \quad (2.28)$$

Las hipótesis de la prueba Durbin-Watson son:

$H_0: \rho = 0$ (Las perturbaciones, u_t , no dependen de sus retrasos de primer orden, u_{t-1} ; no hay autocorrelación de primer orden)

$H_1: \rho \neq 0$ (Las perturbaciones dependen de sus retrasos de primer orden; hay autocorrelación de primer orden)

El estadígrafo de Durbin y Watson, representado usualmente como “d”, se calcula a partir de los residuos del modelo y de sus retardos de primer orden:

$$d = \frac{\sum_{i=2}^n (e_t - e_{t-1})^2}{\sum_{i=1}^n e_i^2}$$

Resulta tan sencilla y es tan divulgada esta prueba, como se mencionaba antes, que muchos softwares reportan directamente en la propia salida de los estimadores y errores estándar de una ecuación de regresión, junto con otros estadígrafos de interés –como R^2 , F o t– el estadígrafo Durbin y Watson.

2.2.4.6 Supuesto de normalidad en las perturbaciones

El supuesto de normalidad en las perturbaciones lo que hace es asumir un comportamiento probabilístico normal para el término de perturbación, o sea:

$u_i \sim N$.

Este supuesto, como se mencionó antes no influye directamente en las propiedades de los estimadores, sino que es requerido para desarrollar los métodos de inferencia utilizados en la validación del modelo. Así, la violación del supuesto hace que pierdan fiabilidad las pruebas F y t, y los intervalos de confianza.

Para analizar la normalidad de una variable existen multitud de pruebas estadísticas, entre ellas la ampliamente usada prueba chi-cuadrado para la bondad de ajuste. Aquí, no obstante, se hará uso casi exclusivo de la prueba Jarque - Bera, relativamente sencilla y con implementación directa en el E-Views.

Prueba Jarque Bera

Las hipótesis de la prueba Jarque Bera, como en general las de las pruebas de este tipo, son:

$$H_0: u \sim N$$

$$H_1: u \not\sim N$$

El estadígrafo de prueba se define como:

$$JB = \frac{n-p}{6} \left[S^2 + \frac{(K-3)^2}{4} \right] \quad (2.22)$$

Este estadígrafo se distribuye bajo la hipótesis nula según una distribución chi-cuadrado con dos grados de libertad. En su cálculo p es la cantidad de parámetros estimados, S es el llamado coeficiente de simetría –mide el grado de simetría de la distribución- y K es el coeficiente de Curtosis –mide el apuntamiento, la elevación de la distribución y la región crítica toma la forma:

$$W = \left\{ JB : JB > X^2_{1-\alpha(2)} \right\} \quad (2.23)$$

Conviene tener presente que si el tamaño de muestra disponible es suficientemente grande el cumplimiento o no de este supuesto deja de ser importante.

2.3 Conclusiones parciales

- 1- Se describe una metodología para obtener los modelos de regresión teniendo en cuenta tres tipos de modelos matemáticos y se utiliza el coeficiente de determinación como medida de calidad del ajuste.
- 2- Se describen varias pruebas estadísticas para analizar la validez de los modelos de regresión y la significación individual de los coeficientes de los modelos. Para comparar y discriminar entre los modelos se aplica el principio de parsimonia
- 3- Se analiza la importancia de realizar el análisis de supuestos que soportan a un modelo de regresión lineal y se describen diversas pruebas con este propósito.

CAPITULO IIIElaboración de modelos matemáticos para la simulación energética de enfriadora de agua helada a flujo constante

El objetivo de este estudio se basa en construir los modelos matemáticos que permitan simular la potencia eléctrica de una enfriadora de agua. Para la construcción y selección del modelo se empleará la metodología descrita en el capítulo 2. Además, se emplearán los datos que ofrece el fabricante que se observan en la tabla 3.1.

Tabla 3.1: Datos de operación de una enfriadora de agua tipo tornillo obtenido del catálogo técnico. Fuente ()								
CAP(kW)	COMP(kW)	Unit(kW)	flujo volumétrico	flujo másico	COP	DEP	Ts	Tconde.in
91	16.6	16.60	15.70	4.36	5.48	19.5	6	30
87.1	18.1	18.10	15.00	4.17	4.81	17.8	6	35
82.8	19.8	19.80	14.30	3.97	4.18	16.1	6	40
78.2	21.7	21.70	13.50	3.75	3.60	14.3	6	45
73.2	23.9	23.90	12.60	3.50	3.06	12.5	6	50
67.8	26.3	26.30	11.70	3.25	2.58	10.8	6	55
94.3	16.8	16.80	16.20	4.50	5.61	20.8	7	30
90.3	18.3	18.30	15.60	4.33	4.93	19.1	7	35
86	20	20.00	14.80	4.11	4.30	17.3	7	40
81.2	21.9	21.90	14.00	3.89	3.71	15.5	7	45
79.2	22.7	22.70	13.60	3.78	3.49	14.7	7	47
76.1	24.1	24.10	13.10	3.64	3.16	13.6	7	50
70.6	26.4	26.40	12.20	3.39	2.67	11.7	7	55
97.6	16.9	16.90	16.80	4.67	5.78	22.3	8	30
93.6	18.4	18.40	16.10	4.47	5.09	20.5	8	35
89.1	20.1	20.10	15.40	4.28	4.43	18.6	8	40
84.3	22	22.00	14.50	4.03	3.83	16.7	8	45
79.1	24.2	24.20	13.60	3.78	3.27	14.6	8	50
73.4	26.6	26.60	12.60	3.50	2.76	12.6	8	55
101	17.1	17.10	17.40	4.83	5.91	23.9	9	30
96.9	18.5	18.50	16.70	4.64	5.24	22	9	35
92.4	20.2	20.20	15.90	4.42	4.57	20	9	40
87.4	22.2	22.20	15.10	4.19	3.94	17.9	9	45
82.1	24.3	24.30	14.10	3.92	3.38	15.8	9	50
76.3	26.7	26.70	13.10	3.64	2.86	13.7	9	55
104.4	17.2	17.20	18.00	5.00	6.07	25.5	10	30
100.2	18.7	18.70	17.30	4.81	5.36	23.5	10	35
95.6	20.4	20.40	16.50	4.58	4.69	21.4	10	40
90.6	22.3	22.30	15.60	4.33	4.06	19.2	10	45
85.1	24.5	24.50	14.70	4.08	3.47	17	10	50
79.3	26.9	26.90	13.70	3.81	2.95	14.7	10	55

107.8	17.3	17.30	18.60	5.17	6.23	27.2	11	30
103.5	18.8	18.80	17.80	4.94	5.51	25.1	11	35
98.9	20.5	20.50	17.00	4.72	4.82	22.9	11	40
93.7	22.4	22.40	16.20	4.50	4.18	20.6	11	45
88.2	24.6	24.60	15.20	4.22	3.59	18.2	11	50
82.2	27	27.00	14.20	3.94	3.04	15.9	11	55

3.1 Selección de las variables

La selección de las variables es el primer paso en la construcción del modelo. Como se explicó al principio se desea determinar la potencia eléctrica de la enfriadora por lo cual se establece la misma como variable predictora.

Se construye la matriz de correlación que se observa en la tabla 3.2 donde se puede apreciar el tipo de relación que existe entre la variable dependiente y las posibles variables independientes

Tabla3.2: Matriz de correlación entre la variable dependiente y las variables independientes								
Variable dependiente		Variables independientes						
		CAP	DEP	Fm	Fv	T1	T2	Tv
Y(pot)	R ²	-0.7990	-0.7927	-0.7984	-0.7980	0.0988	0.9934	0.0619
	Valor P	0.0000	0.0000	0.0000	0.0000	0.5606	0.0000	0.7158

Donde:

Y-Potencia del Chiller;

CAP-Capacidad nominal;

DEP-Presión en el compresor

Fm-Flujo másico en el compresor

Fv-Flujo volumétrico;

T1-Temperatura del retorno;

T2-Temperatura de condensación

Tv-Temperatura del suministro en el evaporador

Esta tabla anterior muestra las correlaciones momento-producto de Pearson, entre cada par de variables. El rango de estos coeficientes de correlación va de -1 a +1, y miden la fuerza de la relación lineal entre las variables. El segundo número en cada bloque de la tabla es un valor-P que prueba la significancia estadística de las correlaciones estimadas. Valor-P abajo de 0.05

indican correlaciones significativamente diferentes de cero, con un nivel de confianza del 95.0%.

Los resultados de la matriz de correlación son los siguientes:

- Existe una fuerte correlación inversa entre la potencia eléctrica y la capacidad nominal, presión, flujo másico y flujo volumétrico.
- Existe una fuerte correlación directa entre la potencia eléctrica y la temperatura del aire acondicionado.
- No existe correlación lineal entre la potencia eléctrica, temperatura de retorno y la temperatura de suministro sabiendo que el valor-P está por encima de 0.05 lo que indica que no hay significancia estadística de las correlaciones estimada.

Teniendo en cuenta estos resultados se escogen para el modelo a construir el conjunto de variables independientes, temperatura del aire a la entrada del condensador y temperatura de retorno del agua helada. Esta última a pesar que no tiene relación lineal directa, si se encuentra implícita en varias variables que si se relacionan directamente como se observa en la expresión 3.1. Por lo que la misma se puede despejar como se muestra en la ecuación 3.2

$$Q(kW) = m * C_p * (tr - ts)$$

$$tr = \left(\frac{Q}{m * C_p} \right) + ts$$

Quedando de esta forma que el modelo de caja negra que describe la potencia eléctrica en función de ambas temperaturas, ver ecuación 3.3

$$POT_i(kW) = f\{t_1 = (T_{ret,i}), t_2 = T_{cond,i}\}$$

3.2 Selección del modelo de caja negra

Se analizan varios modelos matemáticos que se corresponden con el análisis teórico realizado, particularmente asociados a la regresión lineal simple, regresión polinomial y la regresión lineal múltiple y regresión no lineal. Para estas variables se pueden escribir los siguientes modelos:

Tabla 3.3 Modelos de regresión y sus variantes.	
Variante	Modelo
Regresión lineal simple	
1	$Y = x_0 + x_1 * t_1$
2	$Y = x_0 + x_1 * t_2$
Regresión polinomial de orden 2	
3	$Y = x_0 + x_1 * t_1 + x_2 * t_1^2$
4	$Y = x_0 + x_1 * t_2 + x_2 * t_2^2$
Regresión lineal múltiple	
5	$Y = x_0 + x_1 * t_1 + x_2 * t_2$
Regresión no lineal	
6	$Y = x_0 + x_1 * t_2 + x_2 * t_1 + x_3 * t_2 * t_1 + x_4 * t_2^2 + x_5 * t_1^2 + x_6 * t_2^2 * t_1^2$
7	$Y = x_0 + x_1 * t_1 + x_2 * t_2 + x_3 * t_1 * t_2$
8	$Y = x_0 + x_1 * t_2 + x_2 * t_1 + x_3 * t_2 * t_1 + x_4 * t_2^2 + x_5 * t_1^2$
9	$Y = x_0 + x_1 * t_1 * t_2 + x_2 * t_2^2 * t_1^2$
10	$Y = x_0 + x_1 * t_2^2 * t_1^2$

Se analizan diez variantes asociadas a los tipos de regresión, descritos para la regresión lineal y para la regresión polinomial un modelo en términos de la temperatura de salida de agua en el evaporador y otro en función de la temperatura del aire a la entrada del condensador. Para aplicar los modelos de regresión al ajuste de los datos de las mediciones del fenómeno en estudio, se utilizó el software de distribución Statgraphic y Eviews7, paquete estadístico estudiado durante la asignatura “Probabilidades y estadística”. Se muestra un resumen de los resultados obtenidos incluyendo el coeficiente de determinación.

3.3 Obtener el modelo

Se procede a utilizar el método de los mínimos cuadrados (MCO) para calcular los coeficientes de las variables en cada uno de los modelos, haciendo uso de

los programas Statgraphics y Eview. A continuación, se muestran algunos resultados importantes.

3.3.1 Variante uno, Modelo de regresión lineal simple.

Aquí se estudia la relación entre las variables Potencia (Y) y la temperatura de retorno (t_1). La ecuación del modelo ajustado es:

$$Y = 18.9743 + 0.1849 * t_1$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 1 iteraciones, en este punto la suma de cuadrados residual se aproximó al mínimo.

El estadístico R^2 indica que el modelo, explica 0.976957% de la variabilidad en Y. El error estándar del estimado muestra que la desviación estándar de los residuos es 3.35925. El error absoluto medio (MAE) de 2.85201 es el valor promedio de los residuos. El estadístico de Durbin-Watson (DW) prueba los residuos para determinar si hay alguna correlación significativa basada en el orden en que se presentaron en su archivo de datos. Como es apreciable el valor de R^2 , expresado en porcentaje absoluto, no se acerca a 100%, por lo que, este modelo no ofrece un buen ajuste para la potencia y, por ende, debe ser rechazado como modelo explicativo del fenómeno en cuestión.

Si se analiza gráficamente el comportamiento de la variable en un diagrama de dispersión. Ver figura 3.1, se puede notar que el conjunto de datos, están dispersos por todo el primer cuadrante, pero no próximos a la recta esbozada, lo que confirma la hipótesis previa de que estas variables no presentan relación lineal.

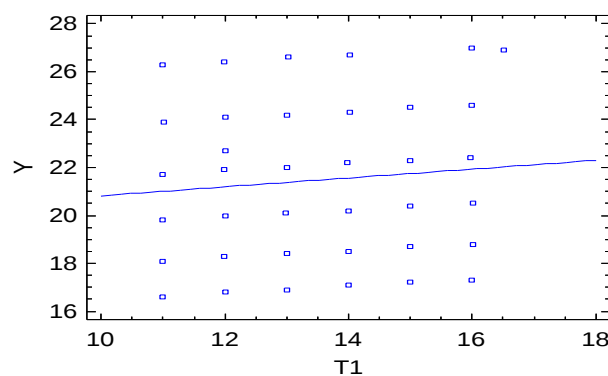


Figura. 3.1: Multicolinealidad entre variable regresora vs variable independiente del modelo 1. Elaboración propia

3.3.2 Variante dos del modelo de regresión lineal simple

En este caso se analiza la relación entre la potencia (Y) y la temperatura de aire acondicionado (t_2).

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 2 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales.

La ecuación del modelo ajustado es

$$Y = 5.0275 + 0.3857 * t_2$$

El estadístico R^2 indica que el modelo explica 98.6842% de la variabilidad en Y. El error estándar del estimado muestra que la desviación estándar de los residuos es 0.387233. Este valor puede utilizarse para construir límites de predicción para nuevas observaciones. El error absoluto medio (MAE) de 0.307067 es el valor promedio de los residuos. El estadístico de Durbin-Watson (DW) prueba los residuos para determinar si hay alguna correlación significativa basada en el orden en que se presentaron en su archivo de datos.

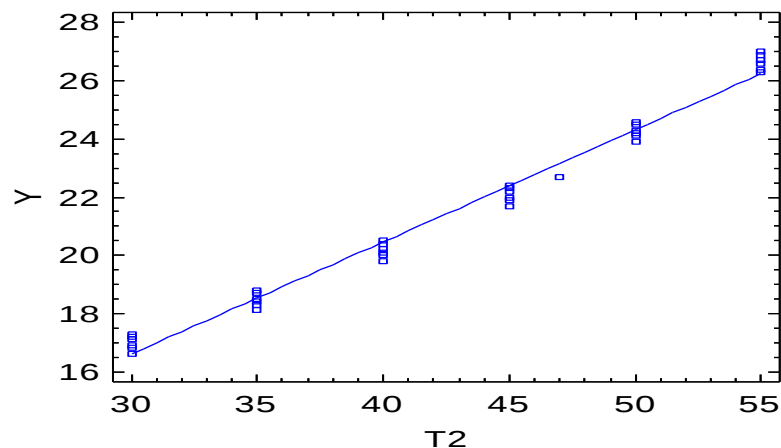


Figura. 3.2: Multicolinealidad entre variable regresora vs variable independiente del modelo 2. Elaboración propia

Este gráfico representa que existe un buen nivel de correlación lineal directa entre la temperatura de condensación (T_2) y la potencia (Y)

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente a } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	5.027534	0.327185	15.36602	0.0000
C(2)	0.385784	0.007530	51.23429	0.0000
R-squared	0.986842	Mean dependent var		21.47027
Adjusted R-squared	0.986466	S.D. dependent var		3.328569
S.E. of regression	0.387233	Akaike info criterion		0.992956
Sum squared resid	5.248217	Schwarz criterion		1.080032
Log likelihood	-16.36968	Hannan-Quinn criter.		1.023654
F-statistic	2624.952	Durbin-Watson stat		0.617177
Prob(F-statistic)	0.000000			

Tabla 3.4 Probabilidad marginal del estadígrafo F (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.4 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, tanto para la fila C (1) como para la fila de C (2), estos valores resultan inferiores al valor de significación prefijado y, por tanto, se rechaza la hipótesis nula, y apoya la significación individual de la temperatura del condensador, para el modelo.

Sobre los supuestos:

- a) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 4.00e-15
Sample Std. Dev. = 0.774029

Method	Value	Probability
t-statistic	3.15E-14	1.0000

Tabla 3.5 Probabilidad marginal del Estadígrafo F (Elaboración propia)

Como se puede apreciar el estadístico t tiene probabilidad marginal de 1, siendo que esta es mayor que el valor de la significación (0,05) no existe

evidencia suficiente para rechazar la hipótesis de nulidad y, por tanto, se cumple el supuesto de media cero en los residuos.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (Hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (Hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.057477	Prob. F(1,35)	0.8119
Obs*R-squared	0.060662	Prob. Chi-Square(1)	0.8055
Scaled explained SS	0.034522	Prob. Chi-Square(1)	0.8526

Tabla 3.6 Probabilidad marginal del Estadígrafo (Elaboración Propia)

Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales superan al valor seleccionado para alfa (0.05) indicando que no se debe rechazar la idea de existencia de homocedasticidad y se cumple este supuesto.

c) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación.

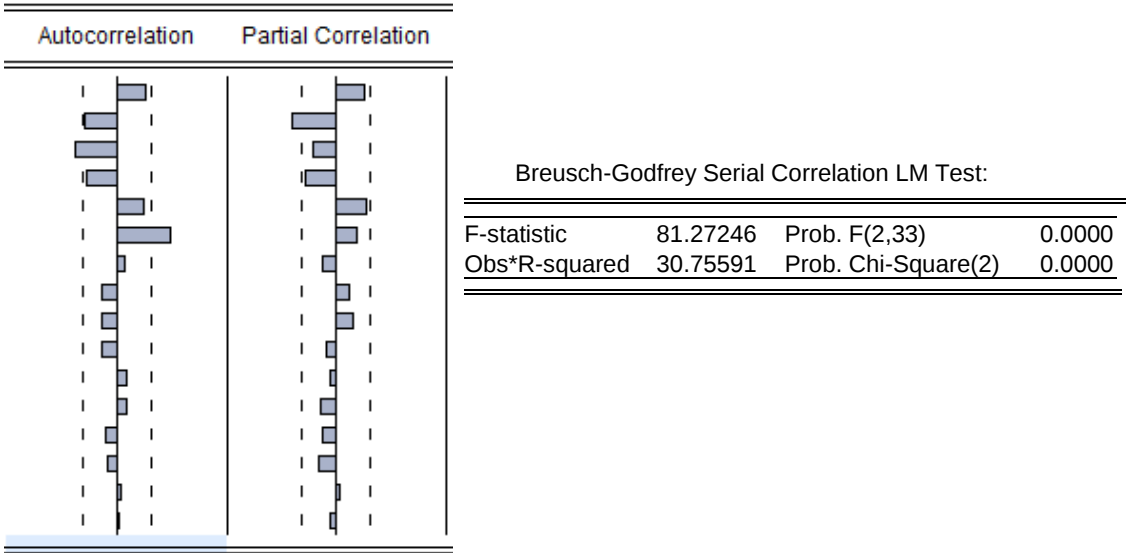


Gráfico 3.1 Autocorrelación y Tablas 3.7 Probabilidades marginales del Estadígrafo (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{No hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{Hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales estas son menores que α (0.05) y se rechaza la hipótesis nula. Se concluye que existe presencia de autocorrelación y se incumple el supuesto.

d) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera.

Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

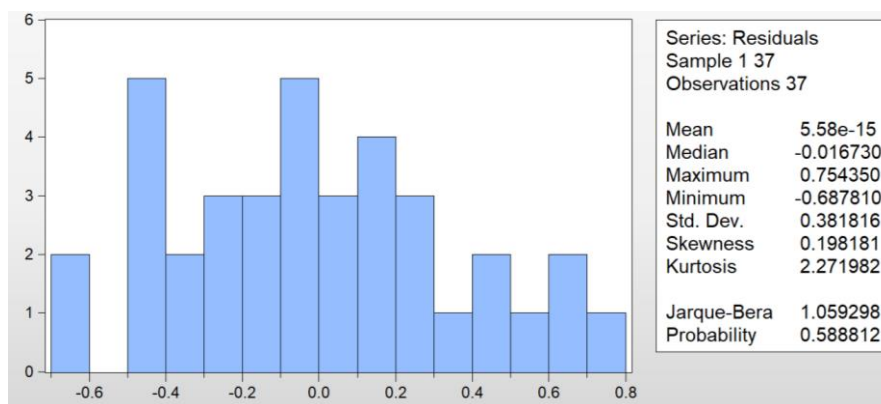


Gráfico 3.2 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

Como se puede apreciar el patrón gráfico 3.1 esboza las barras semejando la campana de Gauss. Además, la probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.588812, superior al 5% que se ha estado trabajando como significación. De aquí que, no se rechace la idea de que los residuos se distribuyen normalmente.

3.3.3 Variante tres, modelo de regresión polinomial de orden 2.

Se probó en este caso un modelo de regresión no lineal para describir la relación entre Y y 1 variables independientes. La ecuación del modelo ajustado es:

$$Y = 41.0274 - 3.11724 * t_1 + 0.1215 * t_1^2$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 5 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales.

El estadístico R^2 indica que el modelo, así ajustado, explica 1.93222% de la variabilidad en Y.

Gráficamente refleja que no existe relación lineal entre las variables, pero también es visible que el conjunto de puntos no se aproxima a la curva de orden dos trazada.

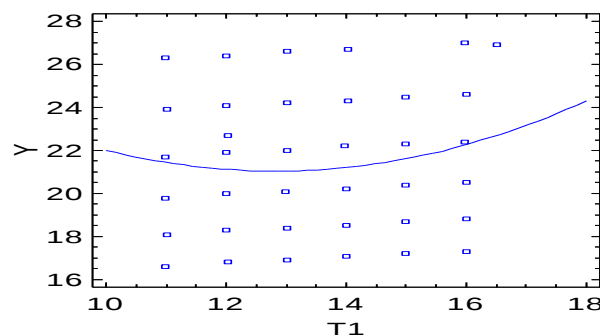


Figura. 3.3: Multicolinealidad entre variable regresora vs variable independiente del modelo 3. (Elaboración propia).

3.3.4 Variante cuatro, modelo de regresión polinomial de orden 2.

Se procede a analizar si la relación entre las variables potencia y temperatura de condensación. La salida muestra los resultados de ajustar un modelo de regresión no lineal para describir la relación entre Y y 1 variables independientes. La ecuación del modelo ajustado es

$$Y = 13.1263 - 0.01148 * t_2 + 0.0046 * t_2^2$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 5 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales.

El estadístico R^2 indica que el modelo, así ajustado, explica 99.4672% de la variabilidad en Y. Este se puede considerar un muy buen ajuste dado que casi ajusta al ciento por ciento de la variabilidad en los datos. El error estándar del estimado muestra que la desviación estándar de los residuos es 0.25001. Este valor puede utilizarse para construir límites de predicción para nuevas observaciones seleccionando la opción de Pronósticos del menú de texto. El error absoluto medio (MAE) de 0.208687 es el valor promedio de los residuos. El estadístico de Durbin-Watson (DW) prueba los residuos para determinar si hay alguna correlación significativa basada en el orden en que se presentaron en su archivo de datos.

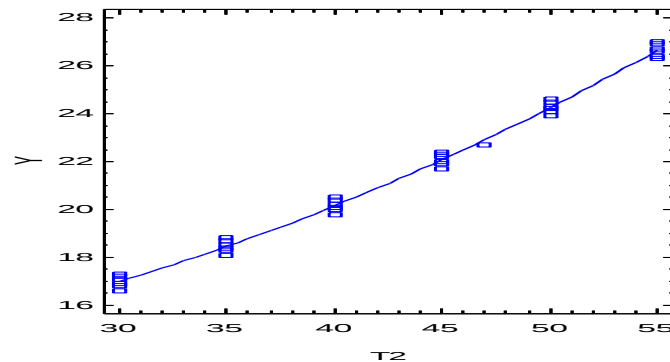


Figura. 3.4: Multicolinealidad entre variable regresora vs variable independiente del modelo 4. Elaboración propia

En la gráfica se puede ver que representa que existe un buen nivel decorrelación directa entre la temperatura del aire.

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	13.12627	1.165047	11.26672	0.0000
C(2)	-0.011484	0.056412	-0.203567	0.8399
C(3)	0.004679	0.000662	7.068582	0.0000
R-squared	0.994672	Mean dependent var		21.47027
Adjusted R-squared	0.994358	S.D. dependent var		3.328569
S.E. of regression	0.250010	Akaike info criterion		0.142972
Sum squared resid	2.125167	Schwarz criterion		0.273587
Log likelihood	0.355021	Hannan-Quinn criter.		0.189020
F-statistic	3173.607	Durbin-Watson stat		0.068196
Prob(F-statistic)	0.000000			

Tabla 3.8 Probabilidad marginal del Estadígrafo(Elaboración Propia)

La probabilidad marginal del estadígrafo F (Prob(F-statistic)) =0.0000 que aparece en la tabla 3.8 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, que para la fila de C (2) no se rechaza la hipótesis de nulidad e indica que el término lineal de esta ecuación se debe retirar del análisis; para la fila de C (3), este valor resulta inferior al valor de significación prefijado y, por tanto, se rechaza la hipótesis nula, y apoya la significación individual de la temperatura del condensador, para el modelo.

Sobre los supuestos:

- a) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 2.88e-15
Sample Std. Dev. = 0.242966

Method	Value	Probability
t-statistic	7.22E-14	1.0000

Tabla 3.9 Probabilidad marginal del estadístico (Elaboración propia)

Como se puede apreciar el estadístico t tiene probabilidad marginal de 1, siendo que esta es mayor que el valor de la significación (0,05) no existe evidencia suficiente para rechazar la hipótesis de nulidad y, por tanto, se cumple el supuesto de media cero en los residuos.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

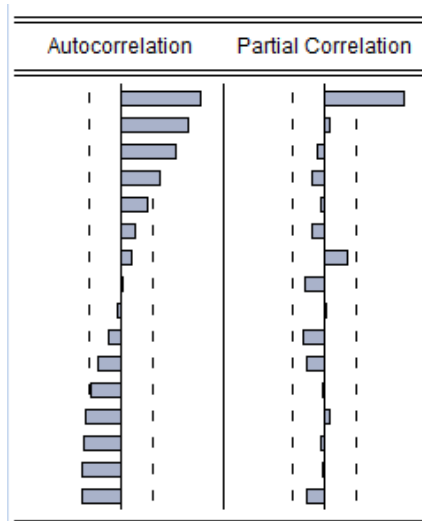
F-statistic	0.015929	Prob. F(1,35)	0.9003
Obs*R-squared	0.016832	Prob. Chi-Square(1)	0.8968
Scaled explained SS	0.005391	Prob. Chi-Square(1)	0.9415

Tabla 3.10 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales superan al valor seleccionado para alfa (0.05) indicando que no se debe rechazar la idea de existencia de homocedasticidad y se cumple este supuesto.

c) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey.



Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	106.4943	Prob. F(2,32)	0.0000
Obs*R-squared	32.16712	Prob. Chi-Square(2)	0.0000

Gráfico 3.3 Autocorrelación y Tablas 3.10 Probabilidad marginal del estadístico (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{no hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales estas son menores que α (0.05) y se rechaza la hipótesis nula. Se incumple el supuesto.

d) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera.

Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

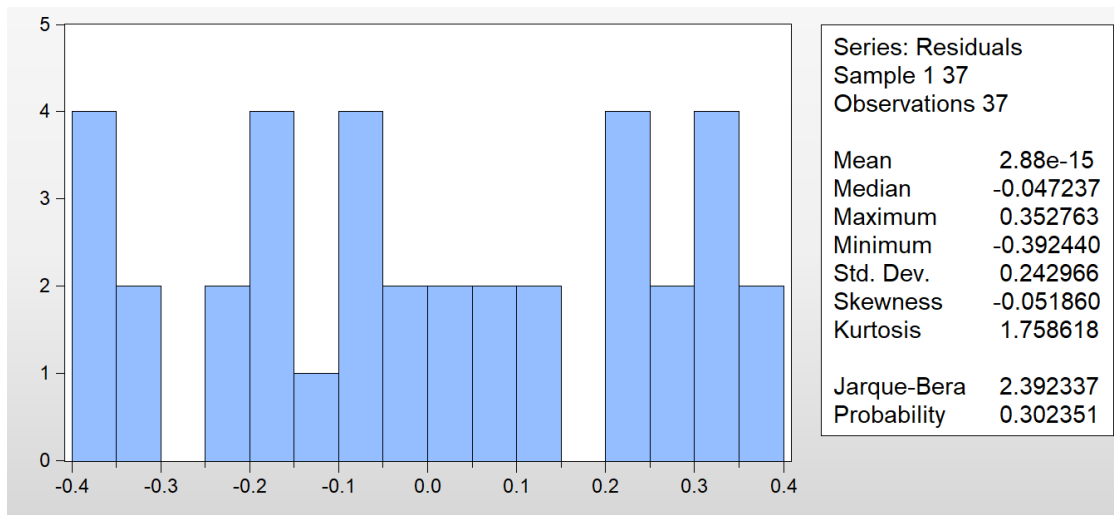


Gráfico 3.4 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

La probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.302351, superior al 5% no se puede rechazar la hipótesis nula y por consiguiente, no se rechaza la idea de que los residuos se distribuyen normalmente.

3.3.5 Variante cinco, modelo de regresión lineal múltiple

Teniendo en cuenta que la relación entre las variables puede estar dado por relaciones subyacentes del fenómeno, se realiza este análisis teniendo en cuenta ambas variables de temperatura.

Al aplicar el procedimiento de regresión lineal múltiple se obtiene la ecuación del modelo ajustado es

$$Y = 3.1209 + 0.1433 * t_1 + 0.3851 * t_2$$

El estadístico R^2 indica que el modelo, así ajustado, explica 99.2711% de la variabilidad en Y. El error estándar del estimado muestra que la desviación estándar de los residuos es 0.292425. El error absoluto medio (MAE) de 0.243502 es el valor promedio de los residuos. El estadístico de Durbin-Watson (DW) prueba los residuos para determinar si hay alguna correlación significativa basada en el orden en que se presentaron en su archivo de datos. Es visible que este modelo ofrece un buen ajuste para explicar la variabilidad en la variable dependiente.

Un análisis gráfico permite observar que la potencia tiene una relación directa con la temperatura de retorno (figura 3.5 izquierda) y está fuertemente relacionada con la temperatura del condensador (figura 3.5 derecha)

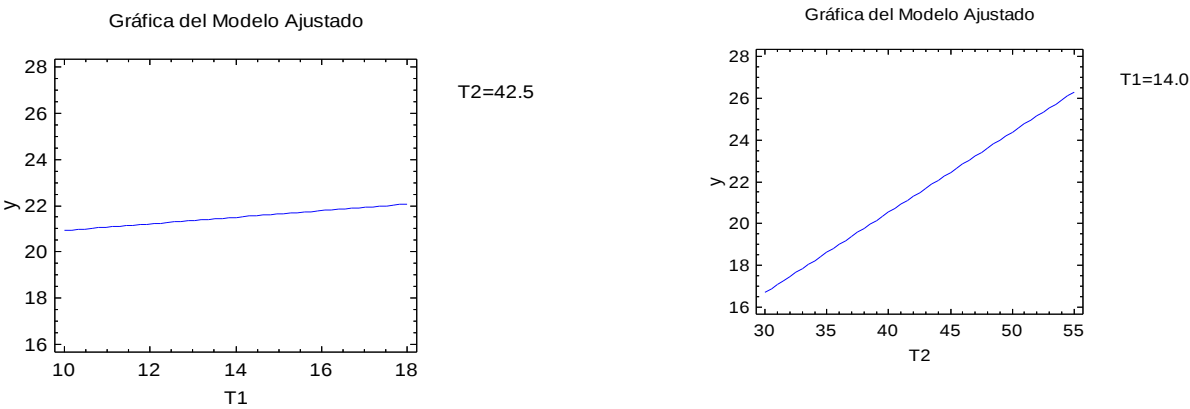


Figura. 3.5: Multicolinealidad entre variable regresora vs variables independientes del modelo 5.
Elaboración propia

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	3.120940	0.440276	7.088598	0.0000
C(2)	0.143348	0.027398	5.231997	0.0000
C(3)	0.385117	0.005688	67.71082	0.0000
R-squared	0.992711	Mean dependent var		21.47027
Adjusted R-squared	0.992282	S.D. dependent var		3.328569
S.E. of regression	0.292425	Akaike info criterion		0.456387
Sum squared resid	2.907420	Schwarz criterion		0.587002
Log likelihood	-5.443166	Hannan-Quinn criter.		0.502435
F-statistic	2315.162	Durbin-Watson stat		1.034268
Prob(F-statistic)	0.000000			

Tabla 3.11 Probabilidad marginal del estadístico (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.11 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, tanto para la fila C(1),C(2) y C(3), estos valores resultan inferiores al valor de significación prefijado y, por tanto, se rechaza la hipótesis nula, y apoya la significación individual de la temperatura del condensador, para el modelo.

Sobre los supuestos:

a) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 5.97e-15

Sample Std. Dev. = 0.284186

<u>Method</u>	<u>Value</u>	<u>Probability</u>
t-statistic	1.28E-13	1.0000

Tabla 3.12 Probabilidad marginal del estadístico (Elaboración propia)

Como se puede apreciar el estadístico t tiene probabilidad marginal de 1, siendo que esta es mayor que el valor de la significación (0,05) no existe evidencia suficiente para rechazar la hipótesis de nulidad y, por tanto, se cumple el supuesto de media cero en los residuos.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.017451	Prob. F(2,34)	0.9827
Obs*R-squared	0.037942	Prob. Chi-Square(2)	0.9812
Scaled explained SS	0.009625	Prob. Chi-Square(2)	0.9952

Tabla 3.13 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales superan al valor seleccionado para alfa (0.05) indicando que no se debe rechazar la idea de existencia de homocedasticidad y se cumple este supuesto.

c) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación

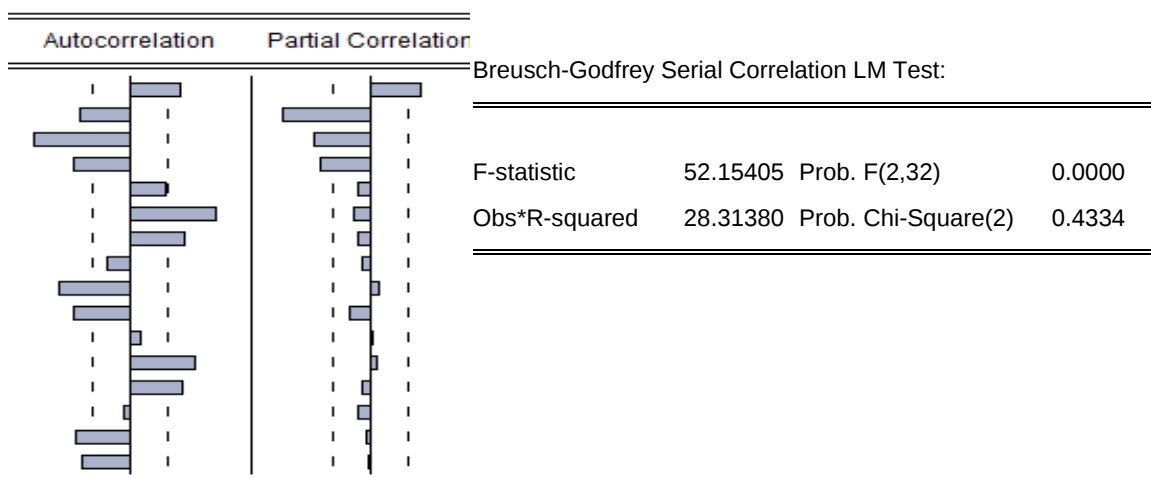


Gráfico 3.5 Autocorrelación y Tabla 3.14 Probabilidad marginal del estadístico (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{no hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales, la segunda de estas es mayor que α (0.05) y no se rechaza la hipótesis nula. Esto indica que se cumple el supuesto.

d) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera. Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

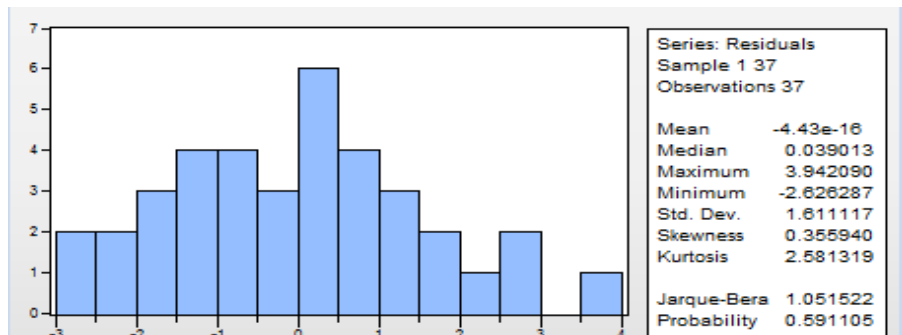


Gráfico 3.6 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

Como se puede apreciar el patrón gráfico esboza las barras semejando la campana de Gauss. Además, la probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.591105, superior al 5% que se ha estado trabajando como significación. De aquí que, no se rechace la idea de que los residuos se distribuyen normalmente.

3.3.6 Variante seis, modelo de regresión no lineal

La salida muestra los resultados de ajustar un modelo de regresión no lineal para describir la relación entre Y y 2 variables independientes.

La ecuación del modelo ajustado es

$$Y = 11.2269 - 0.0379 * t_2 + 0.1595 * t_1 + 0.0029 * t_2 * t_1 + 0.0047 * t_2^2 - 0.0028 * t_1^2 - 0.0000014 * t_2^2 * t_1^2$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 6 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales.

El estadístico R^2 indica que el modelo, así ajustado, explica 99.9907% de la variabilidad en Y. Se hace notar que este modelo proporciona un muy buen ajuste. El error estándar del estimado muestra que la desviación estándar de los residuos es 0.0352159. El error absoluto medio (MAE) de 0.0254245 es el valor promedio de los residuos.

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	11.22688	1.631282	6.882241	0.0000
C(2)	-0.037945	0.049020	-0.774080	0.4449
C(3)	0.159512	0.170414	0.936029	0.3567
C(4)	0.002953	0.003550	0.831889	0.4120
C(5)	0.004782	0.000301	15.87773	0.0000
C(6)	-0.002795	0.003899	-0.716866	0.4790
C(7)	-1.44E-06	1.51E-06	-0.949838	0.3498
R-squared	0.999907	Mean dependent var	21.47027	
Adjusted R-squared	0.999888	S.D. dependent var	3.328569	
S.E. of regression	0.035216	Akaike info criterion	-3.685980	
Sum squared resid	0.037205	Schwarz criterion	-3.381212	
Log likelihood	75.19063	Hannan-Quinn criter.	-3.578535	
F-statistic	53597.92	Durbin-Watson stat	1.817445	
Prob(F-statistic)	0.000000			

Tabla 3.16 Probabilidad marginal del estadístico (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.16 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, que para las filas de C (2),c(3),c(4),c(6) y c(7) no se rechaza la hipótesis de nulidad e indica que estos términos se deben retirar delaecuación,para la fila de C(5), este valor resulta inferior al valor de significación prefijado y, por tanto, se rechaza la hipótesis nula, y apoya la significación individual de la temperatura del condensador, para el modelo.

Sobre los supuestos:

- Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 2.78e-15
Sample Std. Dev. = 0.032148

Method	Value	Probability
t-statistic	5.27E-13	1.0000

Tabla 3.17 Probabilidad marginal del estadístico (Elaboración propia)

Como se puede apreciar el estadístico t tiene probabilidad marginal de 1, siendo que esta es mayor que el valor de la significación (0,05) no existe evidencia suficiente para rechazar la hipótesis de nulidad y, por tanto, se cumple el supuesto de media cero en los residuos.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (Hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	3.947906	Prob. F(2,34)	0.0287
Obs*R-squared	6.973133	Prob. Chi-Square(2)	0.0306
Scaled explained SS	3.796640	Prob. Chi-Square(2)	0.1498

Tabla 3.18 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que sólo una de las probabilidades marginales supera el valorseleccionado para alfa(0.05)indicando que se debe rechazar la idea de existencia de homocedasticidad y no se cumple este supuesto.

c) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas

discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación

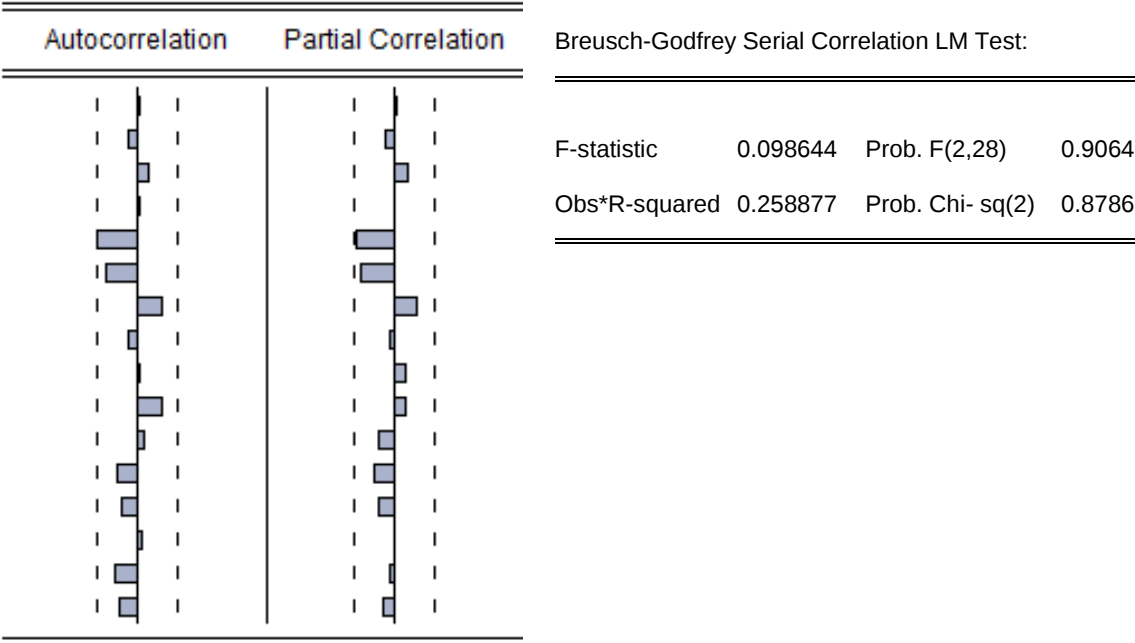


Gráfico 3.7 Autocorrelación y Tabla 3.19 Probabilidad marginal del estadístico (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{no hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales estas son superiores que α (0.05) y no se rechaza la hipótesis nula. Por tanto, se cumple el supuesto

- d) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera. Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

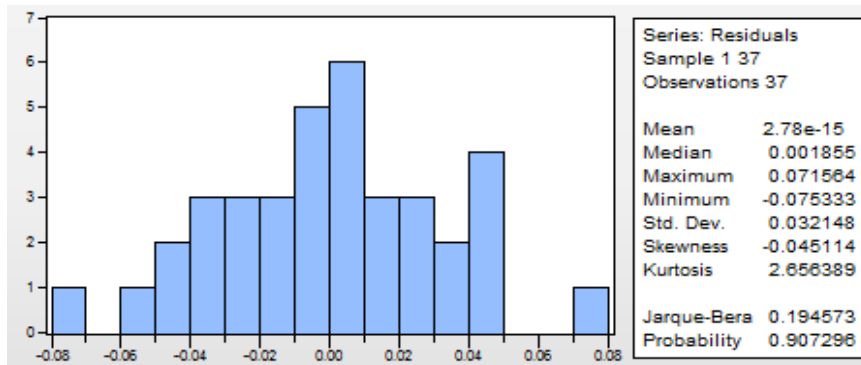


Gráfico 3.8 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

Como se puede apreciar el patrón gráfico esboza las barras semejando la campana de Gauss. Además, la probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.907298, superior al 5% que se ha estado trabajando como significación. De aquí que, no se rechace la idea de que los residuos se distribuyen normalmente.

3.3.7 Variante siete, modelo de regresión no lineal

La salida muestra los resultados de ajustar un modelo de regresión no lineal para describir la relación entre Y y 2 variables independientes. La ecuación del modelo ajustado es

$$Y = 3.3462 + 0.1267 * t_1 + 0.3799 * t_2 + 0.00038 * t_1 * t_2$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 2 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales.

El estadístico R^2 indica que el modelo, así ajustado, explica 99.2714% de la variabilidad en Y. Este estadístico indica que el modelo es adecuado para explicar el comportamiento de la potencia en función de las temperaturas conforme la ecuación descrita anteriormente. El error estándar del estimado muestra que la desviación estándar de los residuos es 0.296758. El error absoluto medio (MAE) de 0.243443 es el valor promedio de los residuos.

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	3.346237	1.926670	1.736799	0.0918
C(2)	0.126738	0.140945	0.899200	0.3751
C(3)	0.379914	0.043671	8.699354	0.0000
C(4)	0.000383	0.003189	0.120213	0.9050
R-squared	0.992714	Mean dependent var		21.47027
Adjusted R-squared	0.992051	S.D. dependent var		3.328569
S.E. of regression	0.296758	Akaike info criterion		0.510004
Sum squared resid	2.906147	Schwarz criterion		0.684157
Log likelihood	-5.435066	Hannan-Quinn criter.		0.571401
F-statistic	1498.707	Durbin-Watson stat		1.036257
Prob(F-statistic)	0.000000			

Tabla 3.20 Probabilidad marginal del estadístico (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.20 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, que para la fila de C(2) y C(4) no se rechaza la hipótesis de nulidad e indica que estos términos de la ecuación se deben retirar, para la fila de C(3), este valor resulta inferior al valor de significación prefijado y, por tanto, se rechaza la hipótesis nula, y apoya la significación individual de la temperatura del condensador, para el modelo.

Sobre los supuestos:

- a) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 4.61e-15
Sample Std. Dev. = 0.284124

Method	Value	Probability
t-statistic	9.87E-14	1.0000

Tabla 3.21 Probabilidad marginal del estadístico (Elaboración propia)

Como se puede apreciar el estadístico t tiene probabilidad marginal de 1, siendo que esta es mayor que el valor de la significación (0,05) no existe evidencia suficiente para rechazar la hipótesis de nulidad y, por tanto, se cumple el supuesto de media cero en los residuos.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.023450	Prob. F(2,34)	0.9768
Obs*R-squared	0.050967	Prob. Chi-Square(2)	0.9748
Scaled explained SS	0.012355	Prob. Chi-Square(2)	0.9938

Tabla 3.22 Probabilidad marginal del estadístico (Elaboración propia)

c) Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales superan al valor seleccionado para alfa (0.05) indicando que no se debe rechazar la idea de existencia de homocedasticidad y se cumple este supuesto.

d) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación.

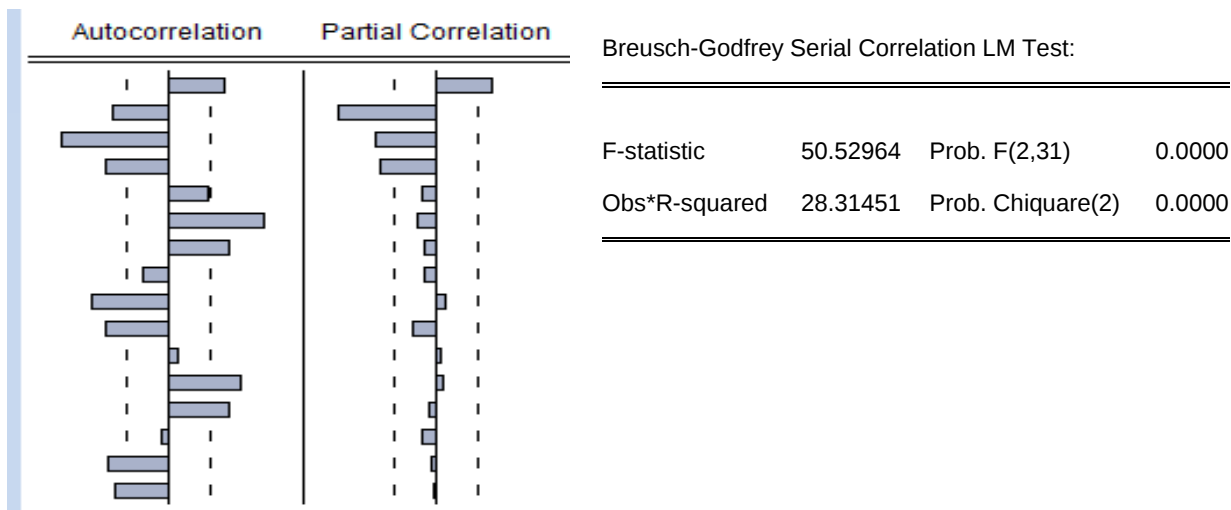


Gráfico 3.9 Autocorrelación y Tabla 3.23 Probabilidad marginal del estadístico (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{no hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales estas no son superiores que α (0.05) y se rechaza la hipótesis nula.

e) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera.

Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

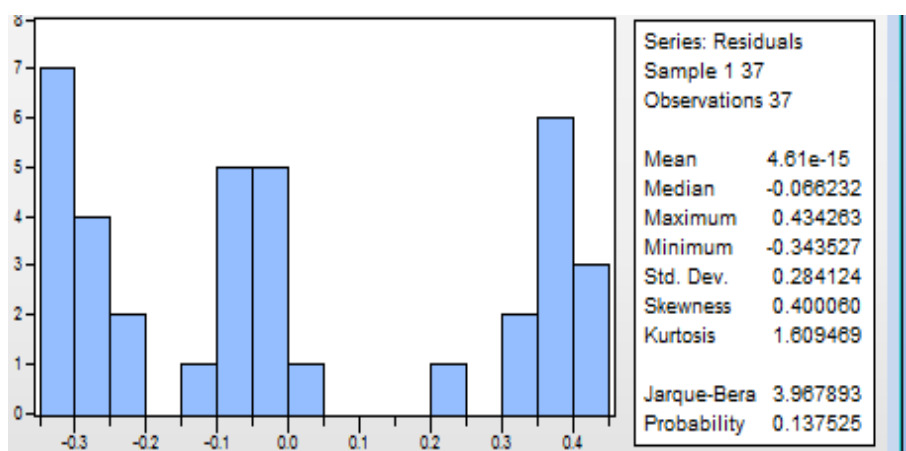


Gráfico 3.10 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

Como se puede apreciar el patrón gráfico esboza las barras semejando la campana de Gauss. Además, la probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.1375, superior al 5% que se ha estado trabajando como significación. De aquí que, no se rechace la idea de que los residuos se distribuyen normalmente.

3.3.8 Variante ocho del modelo no lineal

La ecuación del modelo ajustado es:

$$Y = 9.73872 - 0.0077 * t_2 + 0.3107 * t_1 - 0.00039 * t_2 * t_1 + 0.0045 * t_2^2 - 0.0058 * t_1^2$$

El proceso de estimación terminó exitosamente después de 5 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales. El estadístico R^2 indica que el modelo ajustado, explica 99.9904% de la variabilidad en Y, indicando que este modelo ofrece un buen ajuste. El error estándar del estimado muestra que la desviación estándar de los residuos es 0.0351603. El error absoluto medio (MAE) de 0.0254914 es el valor promedio de los residuos.

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	9.738706	0.453531	21.47308	0.0000
C(2)	0.007758	0.009350	0.829782	0.4130
C(3)	0.310758	0.060615	5.126752	0.0000
C(4)	-0.000399	0.000382	-1.043826	0.3046
C(5)	0.004510	9.37E-05	48.13394	0.0000
C(6)	-0.005829	0.002233	-2.611033	0.0138
R-squared	0.999904	Mean dependent var	21.47027	
Adjusted R-squared	0.999888	S.D. dependent var	3.328569	
S.E. of regression	0.035160	Akaike info criterion	-3.710404	
Sum squared resid	0.038324	Schwarz criterion	-3.449174	
Log likelihood	74.64248	Hannan-Quinn criter.	-3.618308	
F-statistic	64520.90	Durbin-Watson stat	1.831808	
Prob(F-statistic)	0.000000			

Tabla 3.24 Probabilidad marginal del estadístico (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.24 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, que para la fila de C(2) y C(4) no se rechaza la hipótesis de nulidad e indica que ambos términos deben ser retirados del análisis, para la fila de C(3) y C(5), este valor resulta inferior al valor de significación prefijado y, por tanto, se rechaza la hipótesis nula, y apoya la significación individual de la temperatura del retorno y del condensador, para el modelo.

Sobre los supuestos:

a) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 4.13e-15
Sample Std. Dev. = 0.032627

Method	Value	Probability
t-statistic	7.69E-13	1.0000

Tabla 3.25 Probabilidad marginal del estadístico (Elaboración propia)

Como se puede apreciar el estadístico t tiene probabilidad marginal de 1, siendo que esta es mayor que el valor de la significación (0,05) no existe evidencia suficiente para rechazar la hipótesis de nulidad y, por tanto, se cumple el supuesto de media cero en los residuos.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

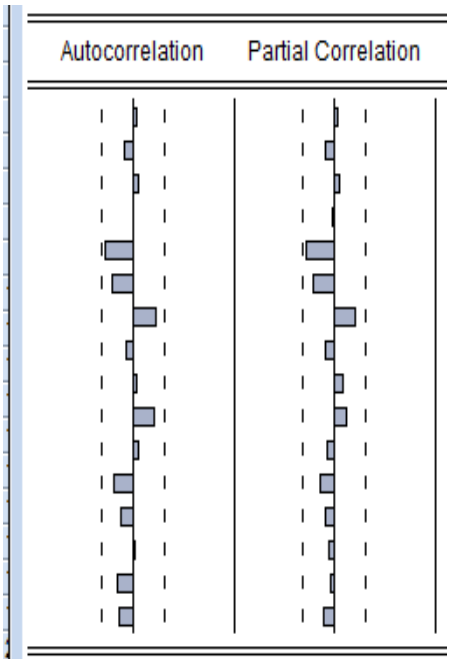
F-statistic	4.304538	Prob. F(2,34)	0.0216
Obs*R-squared	7.475774	Prob. Chi-Square(2)	0.0238
Scaled explained SS	5.234536	Prob. Chi-Square(2)	0.0730

Tabla 3.26 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que sólo una de las probabilidades marginales no supera al valor seleccionado para alfa(0.05) indicando que se debe rechazar la idea de existencia de homocedasticidad y no se cumple este supuesto.

c) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.138879	Prob. F(2,29)	0.8709
Obs*R-squared	0.351019	Prob. Chi-Square(2)	0.8390

Grafico 3.11 Autocorrelación y Tabla 3.27 Probabilidad marginal del estadístico (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0$ no hay autocorrelación.

$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0$ hay autocorrelación

Como se puede apreciar en la columna que menciona las probabilidades marginales estas son superiores que α (0.05) y no se rechaza la hipótesis nula, por tanto, se cumple el supuesto.

d) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera.

Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad

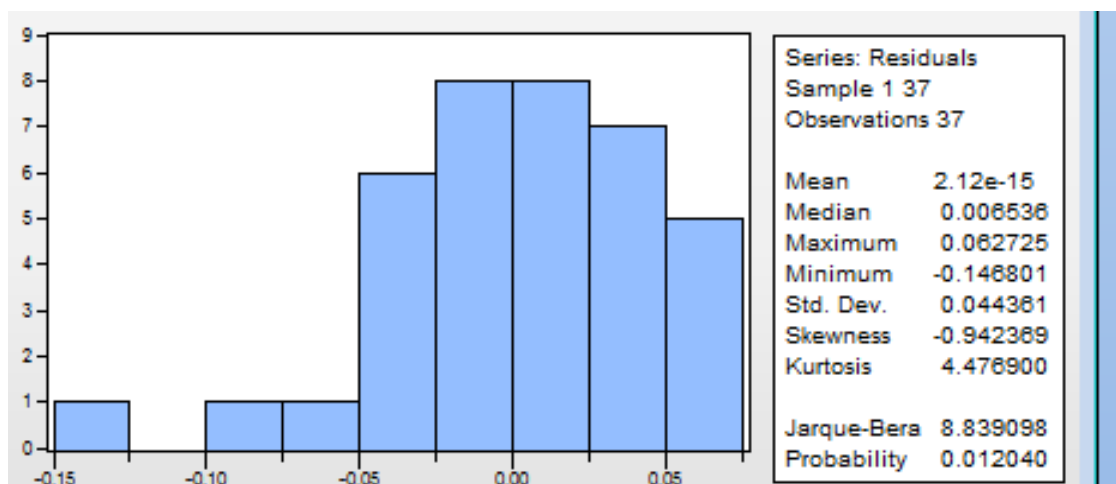


Gráfico 3.12 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

Como se puede apreciar el patrón gráfico esboza las barras semejando la campana de Gauss. Además, la probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.012040, este no es superior al 5% que se ha estado trabajando como significación. De aquí que, se rechace la idea de que los residuos se distribuyen normalmente, y no se cumple el supuesto de normalidad en los residuos.

3.3.9 Variante nueve del modelo matemático

Se ajustan los datos a un modelo de regresión no lineal para describir la relación entre Y y 2 variables independientes. La ecuación del modelo ajustado es:

$$Y = 4.66031 + 0.03813 * t_1 * t_2 - 0.0000014 * t_2^2 * t_1^2$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 6 iteraciones, en este punto la suma de cuadrados residual se aproximó al mínimo.

El estadístico R^2 indica que el modelo, explica el 76.5718% de la variabilidad en Y. El modelo resulta adecuado para explicar la potencia a través de las temperaturas relacionándolas según la ecuación obtenida previamente. El error estándar del estimado muestra que la desviación estándar de los residuos es 1.65783. El error absoluto medio (MAE) de 1.30023 es el valor promedio de los residuos.

Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	4.660042	4.161119	1.119901	0.2706
C(2)	0.038135	0.014118	2.701155	0.0107
C(3)	-1.47E-05	1.15E-05	-1.269300	0.2130
R-squared	0.765718	Mean dependent var		21.47027
Adjusted R-squared	0.751937	S.D. dependent var		3.328569
S.E. of regression	1.657826	Akaike info criterion		3.926496
Sum squared resid	93.44514	Schwarz criterion		4.057111
Log likelihood	-69.64017	Hannan-Quinn criter.		3.972544
F-statistic	55.56208	Durbin-Watson stat		0.658765
Prob(F-statistic)	0.000000			

Tabla 3.28 Probabilidad marginal del estadístico (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.28 contra el valor

de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

También aquí se puede contrastar la significación individual de los coeficientes del modelo para saber si son significativamente diferentes de cero o no, o sea, realizar la prueba enunciada en (2.2.3). Como se puede apreciar en la columna *Prob* de la tabla anterior, que para la fila de C (2) no se rechaza la hipótesis de nulidad e indica que el término de esta ecuación se debe retirar del análisis.

Sobre los supuestos:

a) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = -4.43e-16

Sample Std. Dev. = 1.611117

Method	Value	Probability
t-statistic	-1.67E-15	1.0000

Tabla 3.29 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que la probabilidad marginal supera al valor seleccionado para alfa (0.05) indicando que no se debe rechazar la idea de residuos de media cero y se cumple este supuesto.

b) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.222248	Prob. F(2,34)	0.3072
Obs*R-squared	2.481757	Prob. Chi-Square(2)	0.2891
Scaled explained SS	1.656926	Prob. Chi-Square(2)	0.4367

Tabla 3.30 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales superan al valor seleccionado para alfa (0.05) indicando que no se debe rechazar la idea de existencia de homocedasticidad y se cumple este supuesto.

c) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación.

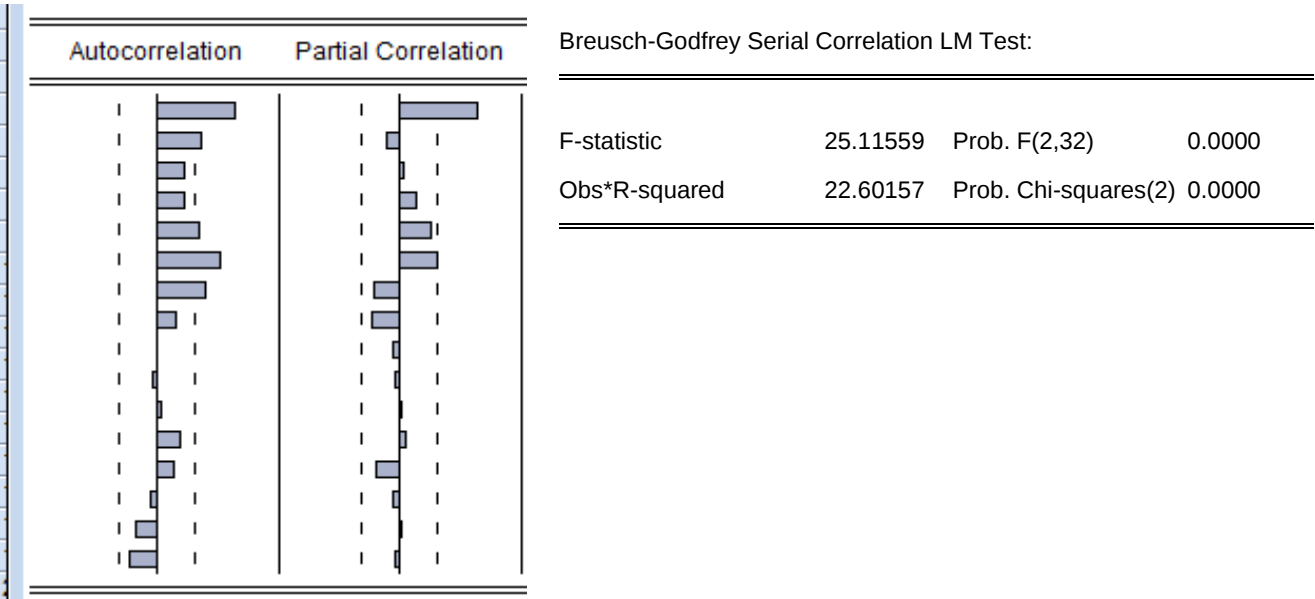


Gráfico 3.13 Autocorrelación y Tabla 3.31 Probabilidad marginal del estadístico (Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{no hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales estas no son superiores que α (0.05) y se rechaza la hipótesis nula.

d) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera.

Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

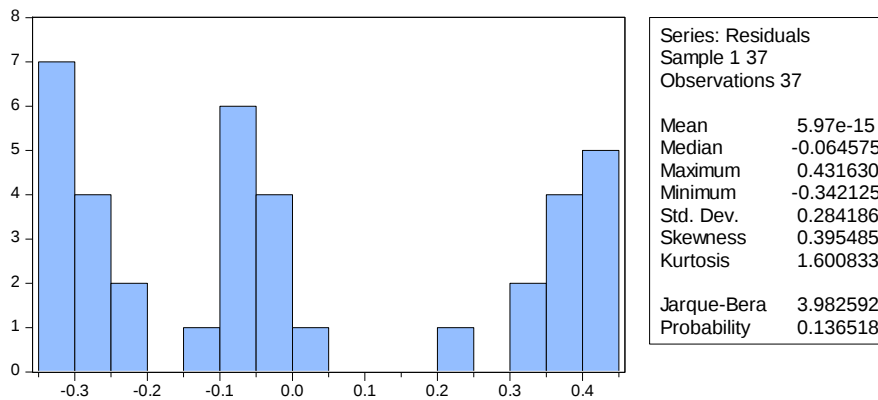


Gráfico 3.14 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

Como se puede apreciar el patrón gráfico no esboza las barras semejando la campana de Gauss. La probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.136518, superior al 5% que se ha estado trabajando como significación. De aquí que, no se rechace la idea de que los residuos se distribuyen normalmente.

3.3.10 Variante diez del modelo matemático

Finalmente, se ajusta un modelo de regresión no lineal que contiene la constante y sólo la combinación cuadrática de los términos variables para describir la relación entre Y y 2 variables independientes.

La ecuación del modelo ajustado es

$$Y = 15.7736 + 0.0000016 * t_2^2 * t_1^2$$

Al realizar el ajuste, el proceso de estimación terminó exitosamente después de 3 iteraciones, en este punto los coeficientes estimados convergieron con los estimados actuales.

El estadístico R^2 indica que el modelo, así ajustado, explica 71.5442% de la variabilidad en Y. El error estándar del estimado muestra que la desviación estándar de los residuos es 1.80078. El error absoluto medio (MAE) de 1.45464 es el valor promedio de los residuos. Para validar la significación del modelo, se aplica la prueba ANOVA, que establece:

$$H_0: \beta_1 = \beta_2 = 0 \text{ frente la } H_1: \exists \beta_i \neq 0,$$

al rechazar la hipótesis nula, se puede concluir a favor de la validez del modelo

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	15.77361	0.675592	23.34782	0.0000
C(2)	1.62E-05	1.73E-06	9.380713	0.0000
R-squared	0.715442	Mean dependent var		21.47027
Adjusted R-squared	0.707312	S.D. dependent var		3.328569
S.E. of regression	1.800778	Akaike info criterion		4.066853
Sum squared resid	113.4980	Schwarz criterion		4.153929
Log likelihood	-73.23678	Hannan-Quinn criter.		4.097551
F-statistic	87.99778	Durbin-Watson stat		0.896605
Prob(F-statistic)	0.000000			

Tabla 3.32 Probabilidad marginal del estadístico (Elaboración propia)

La decisión se puede tomar contrastando el valor de la probabilidad marginal del estadígrafo F (Prob(F-statistic)) que aparece en la tabla 3.32 contra el valor de significación prefijado en $\alpha = 0.05$, siendo este primero menor y, por tanto, se puede rechazar la hipótesis nula. Siendo válido el modelo ajustado.

Como se puede apreciar en la columna Prob. de la tabla anterior, el valor de C (2) es inferior al valor de significación prefijado y, por ende, se rechaza la hipótesis nula y apoya la significación del término que forman las variables, temperatura de retorno y del condensador

Sobre los supuestos:

e) Residuos de media cero.

Se realiza una prueba t para la media de una muestra. Se plantea la hipótesis nula $H_0: \mu_r = 0$ frente la $H_1: \mu_r \neq 0$,

Test of Hypothesis: Mean = 0.000000

Sample Mean = 4.98e-15
Sample Std. Dev. = 1.775591

Method	Value	Probability
t-statistic	1.71E-14	1.0000

Tabla 3.33 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales están por debajo del valor seleccionado para alfa (0.05) indicando

que no se debe rechazar la idea de residuos de media cero y se cumple este supuesto.

f) Homocedasticidad en las perturbaciones

Se realizó el Test: Breusch-Pagan-Godfrey para analizar la presencia de heterocedasticidad.

Se plantea:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (hay homocedasticidad)

$H_1: \text{Al menos un } \alpha_i \neq 0$ (hay heterocedasticidad)

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	4.848154	Prob. F(2,34)	0.0140
Obs*R-squared	8.210383	Prob. Chi-Square(2)	0.0165
Scaled explained SS	6.504902	Prob. Chi-Square(2)	0.0387

Tabla 3.34 Probabilidad marginal del estadístico (Elaboración propia)

Se puede apreciar en la última columna de dicha tabla que las probabilidades marginales no superan al valor seleccionado para alfa (0.05) indicando que se debe rechazar la idea de existencia de homocedasticidad y no se cumple este supuesto.

g) no autocorrelación entre las perturbaciones

Se cumple el supuesto de no autocorrelación de los residuos. Esto se puede apreciar en el correlograma, ningún valor de correlación supera las líneas discontinuas o con la prueba Breusch-Godfrey como alternativa más robusta de la prueba Durbin-Watson de autocorrelación

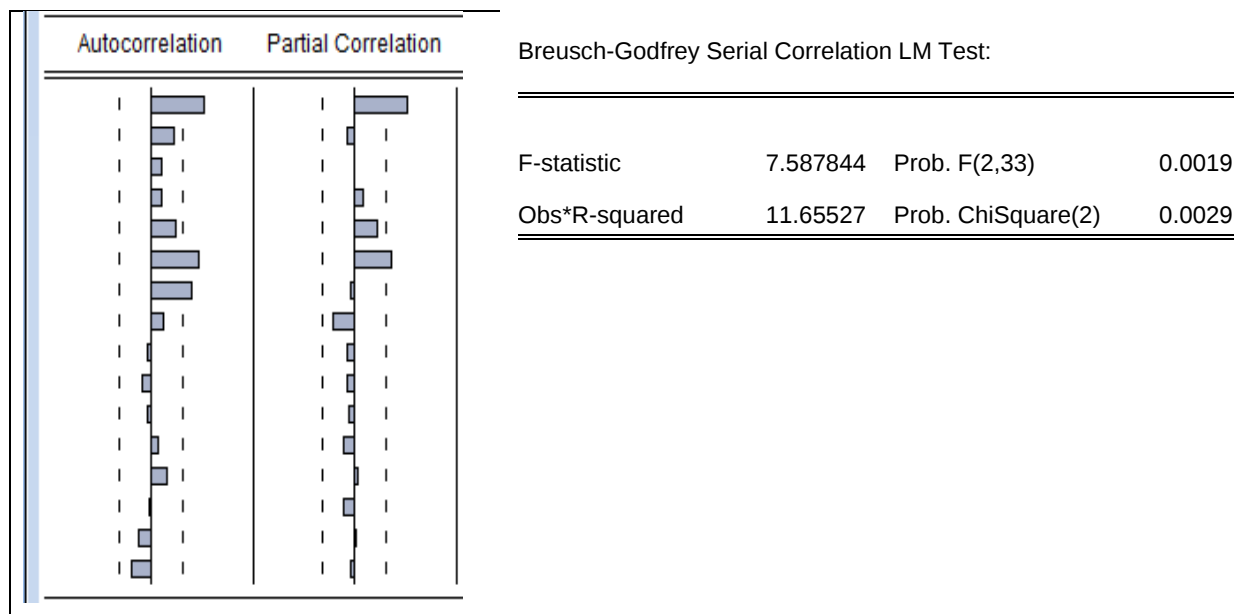


Gráfico 3.15 Autocorrelación y Tabla 3.35 Probabilidad marginal del estadístico
(Elaboración propia)

Las hipótesis de la prueba toman la siguiente forma:

$$H_0: \forall_{i>1} \alpha_i = 0 \text{ y } \forall_j \theta_j = 0 \quad \text{no hay autocorrelación.}$$

$$H_0: \exists_{i>1} \alpha_i \neq 0 \text{ ó } \exists_{i>1} \theta_j \neq 0 \quad \text{hay autocorrelación}$$

Como se puede apreciar en la columna que menciona las probabilidades marginales estas no superan a α (0.05) y se rechaza la hipótesis nula. No se cumple el supuesto

h) normalidad en las perturbaciones

La normalidad de los residuos es probada con el test Jarque-Bera.

Aquí se plantean las hipótesis sobre normalidad en los residuos, por tanto, la hipótesis nula contiene la igualdad respecto a si los residuos se distribuyen normalmente, mientras que la alternativa niega esta posibilidad.

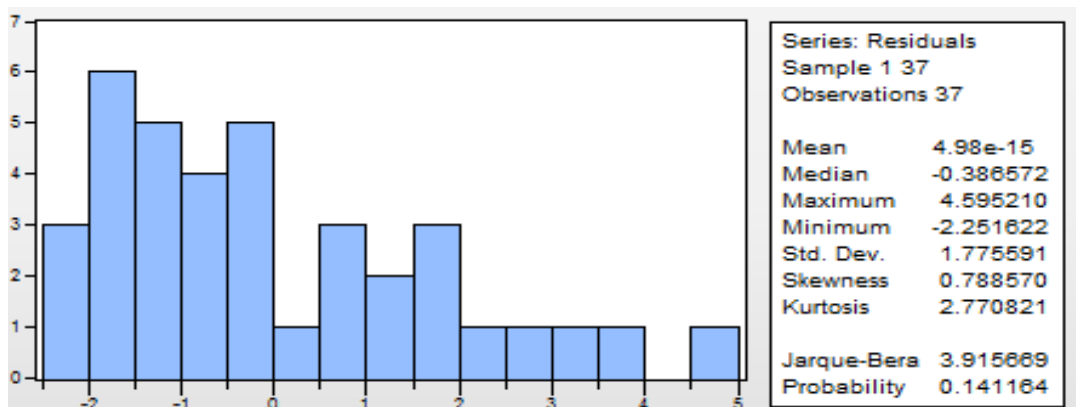


Gráfico 3.14 Probabilidad estadístico Jarque-Bera (Elaboración Propia)

A pesar de que no se aprecia el patrón gráfico esbozado en las barras semejando la campana de Gauss, la probabilidad asociada al estadístico Jarque-Bera (parte inferior derecha) es igual a 0.141164, superior al 5% que se ha estado trabajando como significación. De aquí que, no se rechace la normalidad en los residuos.

3.4. Comparación de modelos

En este trabajo se probaron distintos modelos de los cuales se presenta un resumen de los principales resultados obtenidos.

	Tabla 3.36 Ecuación	R^2 % absoluto
Regresión lineal simple		
1	$Y = 18.9743 + 0.1849 * t_1$	0.97
2	$Y = 5.0275 + 0.3857 * t_2$	98.68
Regresión polinomial de orden 2		
3	$Y = 41.0274 - 3.11724 * t_1 + 0.1215 * t_1^2$	1.93
4	$Y = 13.1263 - 0.01148 * t_2 + 0.0046 * t_2^2$	99.47
Regresión lineal múltiple		
5	$Y = 3.1209 + 0.1433 * t_1 + 0.3851 * t_2$	99.27
Regresión no lineal		
6	$Y = 11.2269 - 0.0379 * t_2 + 0.1595 * t_1 + 0.0029 * t_2 * t_1 + 0.0047 * t_2^2 - 0.0028 * t_1^2 - 0.0000014 * t_2^2 * t_1^2$	99.99
7	$Y = 3.3462 + 0.1267 * t_1 + 0.3799 * t_2 + 0.00038 * t_1 * t_2$	99.27

8	$Y = 9.7388 + 0.0077 * t_2 + 0.3107 * t_1 - 0.00039 * t_2 * t_1 + 0.0045 * t_2^2 - 0.0058 * t_1^2$	99.99
9	$Y = 4.66031 + 0.03813 * t_1 * t_2 - 0.0000014 * t_2^2 * t_1^2$	76.57
10	$Y = 15.7736 + 0.0000016 * t_2^2 * t_1^2$	71.54

Teniendo en cuenta que el estadístico R^2 toma valores de 0 a 100 (expresado en % absoluto), esto indica que algunas regresiones no explican un porcentaje aceptable de la varianza de la variable dependiente, tal es el caso de la variante 1 y la variante 3, y que son desechadas como modelos explicativos. Si, por otra parte, se quiere un modelo con un elevado porcentaje explicativo, fijamos este valor aceptable R^2 por encima del 90%, refinando el proceso de selección del modelo explicativo y retirando del análisis a las variantes 9 y 10. Aunque estos últimos son considerados modelos útiles.

Se presenta a continuación un resumen del comportamiento de los modelos ante las pruebas que validan el cumplimiento de los supuestos exigidos y que fueron realizadas y expuestas previamente.

TABLA 3.37 CUMPLIMIENTO DEL SUPUESTO

VARIANTE	R^2	a)	b)	c)	d)
1	0.98	-	-	-	-
2	98.68	Sí	Sí	No	Sí
3	1.93	-	-	-	-
4	99.47	Sí	Sí	No	Sí
5	99.27	Sí	Sí	Sí	Sí
6	99.99	Sí	No	Sí	Sí
7	99.27	Sí	Sí	No	Sí
8	99.99	Sí	No	Sí	No
9	76.57	Sí	Sí	No	Sí
10	71.54	Sí	No	No	Sí

Se puede notar que la mayoría de los modelos incumplen al menos un supuesto, excepto el modelo de regresión lineal múltiple, el peor caso es aquel modelo que incumpla el supuesto de normalidad de los residuos, dado que este es un requisito inviolable de la regresión. El incumplimiento de otros supuestos

como la no autocorrelación de los residuos puede ser resuelto aplicando algunas medidas remediales que no son objetivo de esta investigación.

Para saber, por otra parte, algunas medidas de calidad de ajuste de los distintos modelos, se resumen a continuación.

MEDIDAS DE CALIDAD			
VARIANTE	R^2	CME	Akaike
1	0.98	-	-
2	98.68	5.25	0.99
3	1.93	-	-
4	99.47	2.13	0.14
5	99.27	2.91	0.46
6	99.99	0.04	-3.69
7	99.27	2.91	0.51
8	99.99	0.04	-3.71
9	76.57	93.45	3.93
10	71.54	113.50	4.07

Como se puede apreciar se presenta además de la variabilidad explicada por cada modelo, el criterio del error cuadrático medio y el criterio de Akaike. En el caso del primero, mientras mayor este sea, mejor es el ajuste. No se comportan igual los demás criterios, en cuyo caso, mientras más cercano a cero mejor es el ajuste que ofrece el modelo. Particularmente, el criterio de Akaike, trata el compromiso de dos funciones, la complejidad del modelo, medido en términos del número de variables explicativas y, la función de ajuste. Puede notarse que los modelos 4, 5 y 7 ofrecen los menores valores de estos indicadores de calidad.

Finalmente, y aplicando el principio de parsimonia de los modelos considerados se debe escoger el más simple. En este caso, se hace notar que el único modelo que validó todos los supuestos es el modelo de regresión lineal múltiple (Variante 5), aunque otros modelos pueden ser usados considerando las limitaciones que presentaron, o sea, el incumplimiento de algún supuesto.

Conclusiones parciales

- 1- Se calculan los modelos planteados, determinando los coeficientes de las variables involucradas, según previamente se había definido en el modelo teórico.
- 2- Se utilizan múltiples pruebas estadísticas para analizar la validez de dichos modelos y el cumplimiento de los supuestos necesarios para una regresión.
- 3- Se calculan algunos indicadores que miden la calidad del ajuste y se comparan los modelos determinándose que el más adecuado es el modelo de regresión lineal múltiple, o sea, la variante 5. También se analiza la posibilidad de uso de otros modelos teniendo en cuenta el supuesto de incumplimiento.

ConclusionesGenerales

1. En la simulación de sistemas energéticos se emplean tres tipos de modelos matemáticos, modelos de caja blanca, negra y gris. De los cuales los más empleados en la ingeniería son los modelos de caja negra debido a las numerosas ventajas que ofrecen.
2. Son escasas las metodologías que permiten la construcción de modelos de caja negra empleando la regresión lineal
3. Se propuso una metodología para la construcción de modelos de caja negra empleando la regresión lineal múltiple que permite la construcción y evaluación de varios modelos simultáneamente.
4. La metodología propuesta es un proceso iterativo y su evaluación se basa en parámetros de calidad de la bondad del ajuste y los supuestos estadísticos
5. Se selecciona el modelo de regresión lineal múltiple (Modelo 5) por presentar un alto nivel de calidad y mínima complejidad de acuerdo con el principio de parsimonia

Recomendaciones

1. Continuar construyendo otros modelos matemáticos para culminar una serie de indicadores que permiten la simulación energética del chiller.
2. Se recomienda construir los diversos tipos de modelos matemáticos que logran una simulación energética completa de un enfriador.
3. Comparar diferentes técnicas de construcciones de modelos de caja negra y verificar su exactitud

Bibliografía

- Aguilera, M. E. (2010). *elementos de econometría para cursos de pregrado .La habana, cuba.*
- Ardakani et al. (2008). A novel approach for optimal chillers loading using particle swarm optimization. *Energy and Buildings*. doi:http://dx.doi:10.1016/j.enbuild.2008.06.010
- Ashrae. (2017). *Fundamentals Handbook*. New york.http://refhub.elsevier.com/S0378-7788%2817%293214-0/Sbref0015
- Astorga Gómez, J. M. (2014). *Aplicación de modelos de regresión lineal para determinar las armónicas de tensión y corriente,Universidad de la Habana,Cuba*
- Beghi,A.Cecchinato,L.Rampazzo,M (2011). A multi-phase genetic algorithm for the efficient management of multi- Chillers systems. *Energy Conversion and Management*. 52, 1650-1661. doi:http://dx.doi:10.1016/j.enconman.2010.10.028
- Beghi, A. C. (2011). A multi-phase genetic algorithm for the efficient management of multi-Chillers systems. *Energy Conversion and Management*. doi:http://dx.doi:10.1016/j.enconman.2010.10.028
- Chan, Tien-Shum.Chang,Yung-Chung.Lu,Jyun-Ting (2015). An analysis of a chiller energy savings using an adaptive neuro-fuzzy inference system. *Energy and Mechanical engineering*.ISBN 978-981-4749-49-7
- Chang, Yung-Chung (2004). a novel energy conservation method—optimal chillers loading. *electric power systems research* 221-226. doi:. http://dx.doi:10.1016/j.epsr.2003.10.012
- Chang, Y. C., C.Y.Lu, J.T.lee, J., & T.S.Chen, C. (2013). Verification of chiller performance promotion and energy saving. *Engineering*, 141-145. doi:http://refhub.elsevier.com/S0378-7788(16)31124-0/sbref0080
- Chang, Y.-C. (2004). *A novel energy conservation method—optimal Chillers loading. Electric Power Systems Research* 69. doi:http://dx.doi:10.1016/j.epsr.2003.10.012
- Chang, Y.-C. (2007). secuencing of chillers by estimating chillers power consumption using artificial neural networks. *building and environment* Vol 42 pp 180-188 doi:. http://dx.doi:10.1016/j.buildenv.2005.08.033
- Chang, Y.-C. L., Lin, F.-A., & Huang., C. (2005). *Optimal Chillers sequencing by branch and boun method for saving energy. Energy Conversion and Management*. doi:doi:10.1016/j.enconman.2004.10.012
- Cheng, Q., & Ch.Wang, S. (2015). Robust optimal design of chiller plant based on cooling load distribution. *Energy Procedia*, 75. doi:https://doi:10.1016/j.egypro.2015.07.212
- Dym, C. (2004). *Principles of mathematical modelling*.Claremont,California: Elsevier.
- Gómez, J. M.-A. (2013). *Aplicación de modelos de regresión lineal para determinar las armónicas de tensión y corriente .universidad de La habana,cuba.*

- Ho, F. Y. (2019). Analysis of a chiller system performance with different componet combination . *Applied thermal Engineering*. 154, 699-710. doi:.
<https://doi.org/10.1016/j.applthermaleng.2019.03.113>
- Jayasekara S. Halgamuge, S. (2014). Optimum sizing and tracking of combined cooling heating and power systems for bulk energy consumers. 124-134.
doi:<http://dx.doi.org/10.1016/j.apenergy.2013.12.040>
- Jiang, W. (2005). framework combining static optimization, dinamic scheduling and decision analysis applicable to complex primary hvac&plants(ph.d. dissertation), drexel university, philadelphia, pa
<http://refhub.elsevier.com/s0378-7788%2814%2900621-5/sbref0055>
- Kaya, D., & Alidrisi, H. (2016). Energy savings potential in air conditioner and chiller systems. *Turkish Journal of Electrical Engineering .Computer Sciences* . 24, 935-945. doi:.
doi:10.3906/elk-1311-204
- Kumbar, A., & N.Pandure, S. (20017). Effect of various parameters on working condition of chiller. *International Conference on Recent Advances in Air Conditioning and Refrigeration* . 109, 479-486. doi:<http://doi:10.1016/j.egypro.2017.03.076>
- Le cam, M., & R.Zmeureanu, y. A. (2013). *Comparison of inverse models used for the forecast of the electric demand of chillers*.France.
- Ledesma, A. H. (2013). Modelado de unidades enfriadoras con fines de optimización. 20-25.
- Li, X., & Wen, j. (2014). Review of building energy modeling for control and operation. *Renewable and sustainable energy reviews*. 37, 517-537.
doi:<http://dx.doi.org/10.1016/j.rser.2014.05.056>
- romero, j.-e., & j.m.belman-flores. (2011). *a simplified black-box model oriented to chilled water temperature control in a variable speed vapour compression system*.
applied thermal engineering 31(2–3):329–35.
- Rossi, T., & J.E.Braun. (1997). *A statistical, rule-based fault detection and diagnostic method for vapor compression air conditioners*. *hvac&r research* 3(1):19–37.
- Santos (2014). Optimal Chillers loading for energy conservation using a new differential cuckoo search approach. 75, 237-243. doi:<http://dx.doi.org/10.1016/j.energy.2014.07.060>
- Seo,B.M.and lee,K.H (2016). Detailed analysis on part load ratio characteristics and cooling energy savings of Chillers staging in a office building. *Energy and Buildings*. 119, 309-322. doi:<http://dx.doi.org/10.1016/j.enbuild.2016.03.067>
- Stanford III, H. W. (2016). *hvac water chillers and cooling towers. fundamentals, application, and operation*. new york: marcel dekker. inc. 270 madison avenue, new york, ny. isbn: 0-8247-0992-6
- Taylor, S. (2012). Optimizing design & control of chiller plant. *ashrae journal*, march 2012.

- Tingting, H., L, C., Xia, B., & Zhihui, F. W. (2019). Study on the feature-recognition-based modeling approach of chillers. *100*, 326-334.
doi:<https://doi.org/10.1016/j.ijrefrig.2019.02.010>
- Underwood, C., & YiK, F. (2004). *Modelling methods for energy in buildings*. blackwell publishing ltd isbn 0-632-05936-2.
- Wang, H. (2017). Empirical model for evaluating power consumption of centrifugal chillers. *Energy and Buildings*. *140*, 359-370. doi:.
<http://dx.doi.org/10.1016/j.enbuild.2017.02.019>
- Wei, X., & G.Kusiak, A. (2014). Modeling and optimizaci3n of a chiller plant, *Energy*. *73*, 898-907. doi:<http://dx.doi.org/10.1016/j.energy.2014.06.102>
- YiK, F., & Lam, V. (1998). Chiller model for plant design studies, *Build*.
doi:[http://refhub.elsevier.com/S0378-7788\(16\)31124-0/sbref0060](http://refhub.elsevier.com/S0378-7788(16)31124-0/sbref0060)
- YU, F., & Chan, K. (2010). . Economic benefits of optimal control for water-cooled chiller systems serving hotels in a subtropical climate. *Energy and Buildings* . *42*, 203-209.
doi:doi:10.1016/j.enbuild.2009.08.016