



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Estadística Aplicada

**MODELO DE MÍNIMOS CUADRADOS APLICADO AL PROCESO DE LA
CRISTALIZACIÓN DEL AZÚCAR**

Ing. Luis Fernando Alvarado Arroyo

Asesorado por el Mtro. Lic. Dorval José Manuel Carías Samayoa

Guatemala, enero de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**MODELO DE MÍNIMOS CUADRADOS APLICADO AL PROCESO DE LA
CRISTALIZACIÓN DEL AZÚCAR**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ING. LUIS FERNANDO ALVARADO ARROYO

ASESORADO POR

EL MTRO. LIC. DORVAL JOSÉ MANUEL CARÍAS SAMAYOA

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, ENERO DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Armando Cruz
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
DIRECTOR	Mtro. Ing. Edgar Darío Álvarez Cotí
EXAMINADOR	Mtro. Ing. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. Ing. William Eduardo Fagiani Cruz
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

MODELO DE MÍNIMOS CUADRADOS APLICADO AL PROCESO DE LA CRISTALIZACIÓN DEL AZÚCAR

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 20 de julio de 2021.

Ing. Luis Fernando Alvarado Arroyo

Facultad de Ingeniería

Decanato
24189101-
24189102
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.015.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **MODELO DE MINIMOS CUADRADOS APLICADO AL PROCESO DE LA CRISTALIZACIÓN DEL AZÚCAR**, presentado por: **Luis Fernando Alvarado Arroyo**, que pertenece al programa de Maestría en artes en Estadística aplicada, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Inga. Aurelia Anabela Cordova Estrada

Decana

Guatemala, enero de 2022

AACE/gaoc



Guatemala, enero de 2022

LNG.EEP.OI.015.2022

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

“MODELO DE MÍNIMOS CUADRADOS APLICADO AL PROCESO DE LA CRISTALIZACIÓN DEL AZÚCAR”

presentado por **Luis Fernando Alvarado Arroyo** correspondiente al programa de **Maestría en artes en Estadística aplicada**; apruebo y autorizo el mismo.

Atentamente,

“Id y Enseñad a Todas”

Mtro. Ing. Edgar Darío Álvarez Cotí
Director

Escuela de Estudios de Postgrado
Facultad de Ingeniería





Guatemala 27 de octubre 2021.

M.A. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Presente

M.A. Ingeniero Álvarez Cotí:

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado **“MODELO DE MÍNIMOS CUADRADOS APLICADO AL PROCESO DE LA CRISTALIZACIÓN DEL AZÚCAR”** de la estudiante **Luis Fernando Alvarado Arroyo** quien se identifica con número de carné **200117406** del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,

MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado

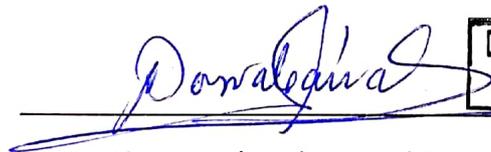
Guatemala, 1 agosto de 2021.

M.A. Ing. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Presente

Estimado M.A. Ing. Álvarez Cotí

Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: **"MODELO DE MÍNIMOS CUADRADOS APLICADO AL PROCESO DE LA CRISTALIZACIÓN DEL AZÚCAR"** del estudiante **Luis Fernando Alvarado Arroyo** del programa de Maestría en **Estadística Aplicada**, identificado con número de carné: **1784 43735 0101**.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.



Lic. Dorval José Manuel Carías Samayoa
Maestro en Desarrollo
Colegiado No. 5958
Asesor de Tesis

ACTO QUE DEDICO A:

Dios	Por darme la fuerza, la sabiduría y la vida.
Mis padres	Arcadio Alvarado y Maria Elena Arroyo, por su motivación y todas sus enseñanzas de vida.
Mi esposa	Silvia Mariana Alonzo, por su amor y apoyo.
Mis hijos	Fernando, Sofía y Ricardo Alvarado, por ser mi inspiración.
Mis hermanos	Anibal, Henry, Daniel, Carlos, Carmen, Reina Alvarado, por estar allí para animarme y motivarme.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por abrir sus puertas al conocimiento.
Facultad de Ingeniería	Por ser la fuente de aprendizaje y experiencias de vida.
Mis docentes de la Facultad	Por compartir sus conocimientos.
Mi asesor	Mtro. Lic. Dorval José Manuel Carías Samayoa, por compartir sus conocimientos y guiarme en la elaboración de este trabajo.
Mis amigos	Ing. Ricardo Meza, Profesor Víctor Priego, por su amistad y apoyo incondicional.
Jefe de área de automatización	Ing. Gerver Cárdenas, por su amistad y por facilitar espacios donde aplicar mis conocimientos.
Compañeras catedráticas	M.A. Licda. Amalia Ruballos, M.A. Licda. Nancy Zurita, por su confianza y apoyo

	2.1.3.3.	Confusión y multicolinealidad	20
2.2.		Cristalización de azúcar	21
3.		PRESENTACIÓN DE RESULTADOS.....	25
3.1.		Identificación de las variables	25
3.2.		Construcción del modelo.....	33
	3.2.1.	Modelo 1.....	33
	3.2.2.	Modelo 2.....	34
	3.2.3.	Modelo 3.....	35
3.3.		Validación del modelo	37
	3.3.1.	Modelo 1.....	38
		3.3.1.1. Bondad de ajuste	38
		3.3.1.2. Prueba de normalidad	38
		3.3.1.3. Prueba de independencia.....	40
		3.3.1.4. Pruebe de homocedasticidad	42
	3.3.2.	Modelo 2.....	44
		3.3.2.1. Bondad de ajuste	44
		3.3.2.2. Prueba de normalidad	45
		3.3.2.3. Prueba de independencia.....	46
		3.3.2.4. Pruebe de homocedasticidad	47
	3.3.3.	Modelo 3.....	48
		3.3.3.1. Bondad de ajuste	48
		3.3.3.2. Prueba de normalidad	48
		3.3.3.3. Prueba de independencia.....	49
		3.3.3.4. Prueba de homocedasticidad	50
	3.3.4.	Resumen.....	51
4.		DISCUSIÓN DE RESULTADOS.....	53

CONCLUSIONES 59
RECOMENDACIONES 61
REFERENCIAS 63
APÉNDICES 67
ANEXOS 69

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Supuestos acerca del error ϵ	13
2.	Gráfica de matriz de todas las variables	27
3.	Efectos principales para la variable de salida Brix	28
4.	Dispersión Brix contra fase compensada	36
5.	Gráfico de probabilidad normal del modelo 1	39
6.	Test de Lilliefors con código de Rstudio.....	40
7.	Residuos contra orden del modelo 1	41
8.	Residuos contra valor ajustado del modelo 1	43
9.	Gráfica de probabilidad normal del modelo 2	45
10.	Residuos contra orden del modelo 2	46
11.	Residuos contra valor ajustado modelo 2	47
12.	Gráfico de probabilidad normal para el modelo 1	49
13.	Residuos contra orden del modelo 3	50
14.	Residuos contra valor ajustado del modelo 3	51

TABLAS

I.	Definición de las variables regresoras y de respuesta	XX
II.	<i>ANOVA para la prueba F</i>	16
III.	Promedio de las variables en cada etapa de la templa.	26
IV.	Análisis de varianza para tacho de refinería	28
V.	Coefficientes para la construcción del modelo de regresión	30
VI.	Resumen del coeficiente de determinación R^2	31

VII.	Coeficientes sin la variable fase compuesta	31
VIII.	Coeficientes de fase y atenuación	32
IX.	Coeficientes de fase compuesta y atenuación.....	32
X.	Combinación de variables regresoras para cada modelo	33
XI.	Análisis de varianza para el modelo de regresión.....	34
XII.	Análisis de Varianza para el modelo 2.....	35
XIII.	Coeficientes para la fase compuesta del modelo cuadrático	36
XIV.	Análisis de varianza del modelo de orden dos.....	37
XV.	Prueba Durbin-Watson modelo 1	42
XVI.	Resultado de prueba Breush-Pagan modelo 1	44
XVII.	Resultados de las pruebas de bondad de ajuste y supuestos	52

LISTA DE SÍMBOLOS

Símbolo	Significado
r	Coefficiente de correlación
R^2	Coefficiente de determinación en regresión múltiple
ϵ	Error del modelo
Z	Estadístico de la distribución normal
β_0	Intersección de la recta
E	Margen de error
n	Muestra
α	Nivel de significancia
β_1	Pendiente de la recta
F	Prueba F o de Fisher
t	Prueba T o de <i>Student</i>
y	Variable de respuesta en el modelo
x	Variable regresora y predictora del modelo

GLOSARIO

ANOVA	Análisis de varianza que se utiliza para encontrar los coeficientes con el método de mínimos cuadrados.
Brix	Media de la concentración de sacarosa en la masa de azúcar.
ECM	Error cuadrático medio.
FIV	Factor de inflación de la varianza.
Homocedasticidad	Término que se refiere a la característica de que las varianzas de los errores son iguales.
Infostat	Programa de computación para análisis estadístico con versión gratuita para estudiantes.
Min	Mínimo entre dos valores.
Minitab	Programa de computación estadístico con gran potencia de herramientas de regresión y análisis.
Promtec	Marca comercial del equipo microondas para medición de la concentración de sacarosa.
Regresión	Proceso estadístico que establece la relación entre variables a través de una ecuación.

RStudio	Programa de computación gratuito que permite realizar modelos y análisis estadístico utilizando código de programación.
Sacarosa	Azúcar que se encuentra en el jugo de la caña dulce la cual se extrae y procesa en un ingenio azucarero.
SCE	Suma de cuadrados de los errores.
SCR	Suma de cuadrados debido a la regresión.
Semillamiento	Proceso de agregar precursores que generan los cristales de azúcar.
Software	Programa de computación para aplicaciones estadísticas.
STC	Suma total de cuadrados.
Tacho	Recipiente donde se generan los cristales de azúcar, trabaja con presión de vacío.
Templa	Etapas de cristalización de un lote de azúcar.

RESUMEN

El objetivo de la presente investigación fue construir un modelo estadístico que relacione las variables en el proceso de cristalización, utilizando los criterios del análisis de regresión y así mejorar la calidad en el tamaño y consistencia del grano de azúcar.

La metodología se centra en la línea de investigación de pronósticos ya que se encontró una ecuación que relaciona las variables del proceso de la cristalización de azúcar tomando la concentración de sacarosa como parámetro de medida. El enfoque del estudio es cuantitativo con un diseño no experimental.

La construcción del modelo se llevó a cabo por medio de la regresión lineal múltiple de primer y segundo orden, con el método de mínimos cuadrados. Se verificaron los componentes del modelo con un análisis de la varianza utilizando software estadístico, interpretando y validado los resultados con pruebas de hipótesis sobre los supuestos de cada modelo o ecuación encontrada.

Se concluyó que el modelo construido, que cumple con los criterios de bondad de ajuste, es de segundo orden, por lo que se estableció que hay una variable que tiene mayor influencia dentro del proceso en la cristalización. Se recomienda validar los supuestos del error para cualquier modelo encontrado antes de parametrizar el equipo de medición.

Los resultados del estudio beneficiarán al proceso de calibración del equipo de medición en los tachos de refinería al proporcionar la metodología de creación

y validación de los supuestos del modelo con un nivel de confianza que asegure la calidad del grano de azúcar en la cristalización.

PLANTEAMIENTO DEL PROBLEMA

- Contexto general

La cristalización por semillado tiene el objetivo de producir la cantidad adecuada de cristales y desarrollarlos homogéneamente para producir masa de tercera. Esta etapa se realiza siguiendo la descripción del proceso del instructivo de operación del tacho, quedando a criterio, experiencia y responsabilidad del operador la calidad en el producto.

La producción de azúcar se hace en un proceso sin pausas en la época de zafra lo que implica que en cada etapa hay distinto operador por turno, lo que genera variación en los criterios de producción afectando la calidad del grano de azúcar.

- Descripción del problema

El proceso de semillado que se realiza en el punto de saturación es determinado por la experiencia del operador con la información de los grados Brix de la muestra enviada previamente al laboratorio y el punto que determine el medidor de Brix. La medición de Brix en el laboratorio se hace con muestras que son extraídas en distintos momentos del punto de saturación y comparándolas con las mostradas por el sistema automatizado.

La decisión sobre el punto óptimo para la cristalización depende de tres factores: experiencia del operador, resultados de laboratorio y comparación con el medidor automático, esto condiciona el proceso y afecta la calidad y pureza

del grano de azúcar. No se conoce un modelo que unifique los criterios tomados por el operario que se compare con la función parametrizada en el controlador lógico programable (PLC), que pudiera controlar el proceso, este modelo ayudaría a unificar los criterios y así ajustar el sistema automático de tal forma que no haya variabilidad de forma manual o automática.

El conocimiento de este modelo haría que el proceso dependiera menos de la experiencia del tachero dejando únicamente que el operario supervise, y en caso de falla del sistema, pueda a partir del modelo encontrado continuar de forma manual.

- Formulación del problema
 - Pregunta central

¿Cuál es el modelo de regresión que relaciona significativamente las variables de saturación, alimentación de materiales, presión de vapor, temperatura y nivel de masa que intervienen en la cristalización de los granos de azúcar?
 - Preguntas auxiliares
 - ¿Cómo influyen las distintas variables de saturación, alimentación de materiales, presión de vapor, temperatura, nivel de masa, fase atenuación y grados Brix en el tamaño y consistencia del grano de azúcar en el proceso de cristalización?

- ¿Qué modelo de regresión simple o multivariable es el adecuado para la optimización del proceso de cristalización?
 - ¿Qué pruebas y supuestos estadísticos se deben realizar para validar un modelo con el que se pueda hacer predicciones para encontrar el punto óptimo de la cristalización del azúcar?
- Delimitación del problema

El problema se analizó para tachos de refinería utilizando una base de datos la zafra de la temporada 2020-2021, tomados del dispositivo de automatización, los datos de calibración y registros manuales del proceso llevado a cabo por el tachero en la elaboración de masa de tercera por semillado.

La población objeto de estudio es la saturación de sacarosa de cada templa de cristalización de azúcar, la base de datos proviene de los registros de calibración para el control de calidad del proceso.

OBJETIVOS

- General

Construir un modelo estadístico que relacione las variables que intervienen en el proceso de cristalización, utilizando los criterios del análisis de regresión y así mejorar la calidad en el tamaño y consistencia del grano de azúcar.

- Específicos

- Identificar las variables que influyen en el proceso de cristalización del azúcar, utilizando metodología de análisis de varianza.
- Construir el modelo que mejor se ajusta al proceso interno programado en el sistema automático por medio del análisis de regresión.
- Validar el modelo propuesto por medio de pruebas de hipótesis que verifiquen los supuestos de la ecuación obtenida.

RESUMEN DEL MARCO METODOLÓGICO

A continuación, se describen las características, la unidad de análisis y las fases del estudio, los cuales fueron tomados en cuenta para desarrollar los objetivos planteados

- Características del estudio

El enfoque del estudio realizado es cuantitativo, ya que se analizaron los valores numéricos continuos de las variables que intervienen en la elaboración de cristales de azúcar.

El alcance es descriptivo correlacional dado que se asociaron los valores de las variables que intervienen en el proceso para establecer una correlación entre ellos y por medio de pruebas estadísticas y análisis de correlación multivariada, se determinó el nivel de asociación existente entre ellas.

El diseño adoptado fue no experimental dado que se analizaron los datos históricos del proceso de cristalización de azúcar en su estado original sin ninguna manipulación; además será transversal pues se estudiará la relación puntual de las variables en cada templa de elaboración de azúcar.

- Unidades de análisis

La población en estudio fue la concentración de sacarosa formada por las mediciones de los parámetros de las variables de cada ciclo o templa de cristalización de azúcar para obtener la pureza y calidad de los granos, de la cual

se extraerán muestras de forma aleatoria simple, que serán estudiadas en su totalidad.

- Variables

A continuación, se presenta las variables analizadas para la creación del modelo.

Tabla I. **Definición de las variables regresoras y de respuesta**

Variable	Definición teórica	Definición operativa	Escala	Símbolo
Concentración de sacarosa (° Brix)	Cociente total de materia seca generalmente azúcares disueltos en un líquido.	Brix	Razón	B _x
Atenuación (dB)	Medida de la recepción de señal entre la antena emisora y receptora dentro del tacho.	Atten	Intervalo	A _t
Temperatura (°)	Nivel térmico en la masa y en el agua dentro del tacho.	Temp	Intervalo	T _m
Fase (°)	Es el movimiento horizontal de la onda transmitida del emisor al receptor, se origina por la distancia que recorre y el objeto que cruza.	Phase	Razón	f
Fase compensada (°)	Es la fase compensada con la temperatura del proceso, la temperatura del medio y un coeficiente de temperatura.	Phase Comp	Razón	f _c

Fuente: elaboración propia.

- Fases del estudio

Se realizaron cinco fases de estudio que se presentan a continuación.

- Fase 1: revisión bibliográfica

Se revisó toda la información bibliográfica necesaria para realizar el estudio, desde el análisis de la investigación previa sobre la automatización del proceso de cristalización de azúcar hasta la búsqueda de herramientas estadísticas útiles para encontrar el grado de asociación y correlación entre las distintas variables para la creación del modelo que mejor ajuste al proceso.

La investigación sobre los procesos de cristalización de granos de azúcar para definir las variables que intervienen en el control de la concentración y los parámetros con los que debe operar.

- Fase 2: gestión de la información

Se realizó una visita a un ingenio azucarero para observar el proceso de toma de muestras para la calibración del equipo de medición de la concentración de sacarosa. Este proceso lo realizan cuando el equipo de medición muestra un margen de error superior a 1 °Brix en comparación con las mediciones manuales que realizan en el laboratorio de la fábrica. Antes de modificar los parámetros del aparato hacen un muestreo de varias plantas para confirmar un desfase en las mediciones.

De los datos históricos provenientes de calibración de las zafas anteriores se calculó la muestra necesaria para un nivel de confianza del 95 %. A pesar de tener un número de muestra adecuado el protocolo de calibración estipula 24

datos de cada variable en las distintas fases del proceso los cuales se utilizaron para todo el análisis estadístico.

- Fase 3: análisis de la información

Las bases de datos de las variables que intervienen en el proceso se analizaron de forma gráfica en busca de una relación y asociación y así aplicar técnicas y pruebas que muestren la influencia que tienen en el proceso.

Al tener el panorama de las variables más influyentes en la medición de la saturación, se formularon combinaciones entre ellas buscando que la significancia y el factor de inflación de la varianza fueran adecuados.

- Fase 4: interpretación de información

Al encontrar las variables que intervienen directamente en la cristalización se definieron los modelos de regresión que mejor se ajusten al proceso. Para ello se probará modelos con transformaciones, de regresión lineal múltiple de primer y segundo orden.

Con los modelos definidos al cumplir con las pruebas de linealidad, coeficiente de determinación aceptable, contrastes validados con pruebas de significancia y un coeficiente de inflación de la varianza apropiado, se procedió a validar de forma gráfica los supuestos del error del modelo aplicando las pruebas de homocedasticidad, normalidad e independencia. El mejor modelo es el que cumpla todos los supuestos y la parsimonia estadística.

- Fase 5: redacción de informe final

Con los resultados obtenidos se procedió a la elaboración del informe desarrollado por objetivos y resumidos con tablas y gráficos, la discusión de resultados llevó a la elaboración de conclusiones y recomendaciones.

INTRODUCCIÓN

El presente estudio trata de una sistematización al construir un modelo de regresión que incluya las variables que influyen en la medición de la concentración de sacarosa de las etapas de cristalización de azúcar, haciendo uso del planteamiento de modelos de regresión lineal múltiple para obtener la ecuación que mejor representa este proceso.

El problema en la cristalización de azúcar se da porque a pesar de tener dispositivos de medición por microondas de alta tecnología que definen cada etapa, un operador sin tanta experiencia no puede realizar el proceso dado que no se tiene una calibración exacta de estos equipos, por lo que se necesitan de una ecuación validada estadísticamente que relacione las variables de medición.

El principal aporte de esta investigación es desarrollar modelos de regresión utilizando el método de mínimos cuadrados con el que se encuentran los coeficientes de las variables regresoras, el análisis de varianza que proporciona información para validar la bondad de ajuste y probar los supuestos del modelo, con lo que se obtiene una ecuación validada estadísticamente con un nivel de confianza que asegura una parametrización adecuada para los instrumentos de medición en el proceso de cristalización.

El estudio tiene un enfoque cuantitativo no experimental al analizar los valores numéricos de las variables, tomados del muestreo realizado para la calibración del equipo de medición de la concentración de sacarosa, el cual tiene un alcance correlacional por la asociación de las variables.

Se desarrollaron distintas fases como la revisión bibliográfica donde se investiga sobre la elaboración de cristales de azúcar, las mediciones que conlleva el proceso y los métodos estadísticos necesario para dar solución al problema. En la fase de la gestión y análisis de la información se cuenta con la colaboración de un ingenio de la costa sur que brinda acceso a sus datos de producción y a la documentación técnica del equipo de medición, lo que hace factible la construcción del modelo a través del análisis de la información.

Se desarrolló una fase de interpretación y discusión de los resultados obtenidos, en donde se identifican las variables influyentes y se concluyó que el mejor modelo de ajuste sobre las variables que influyen en la medición de la concentración y por ende en la cristalización de azúcar es de orden dos, de una sola variable.

El informe final del estudio está estructurado en cuatro capítulos que se discuten a continuación: en el primer capítulo, se presentan los estudios previos sobre el proceso de la elaboración de cristales de azúcar en los aparatos de cocción, las etapas o templeas que se deben controlar desde la concentración y saturación de sacarosa y la revisión de los sistemas de control y de medición que son el tema principal en la calidad de la producción. Asimismo se presentan la metodología estadística que se utiliza para dar solución a este tipo de problemática.

En el segundo capítulo, se desarrolló el marco teórico con la información literaria necesaria para describir y analizar los datos, definir los modelos de mejor ajuste con las ecuaciones que sugiere la metodología de mínimos cuadrados y la pruebas para validar un modelo. En el tercer capítulo, se presentan los resultados obtenidos desarrollados sobre los objetivos planteados, se elaboran gráficas y tablas que sintetizan la información del análisis de

varianza, se interpretan los resultados y se formulan las ecuaciones con las variables estadísticamente significativas del proceso.

En el cuarto capítulo, se discutió los resultados, en donde se hace un análisis interno y externo de la investigación que permite determinar un modelo con el máximo nivel de ajuste, esta discusión ayuda a discernir sobre el alcance los objetivos y la solución del problema planteado.

1. MARCO REFERENCIAL

1.1. Estudios previos

El proceso de la producción de azúcar ha incorporado diferentes tecnologías para su elaboración, sin embargo, conserva algunos procesos heredados del procesamiento artesanal clásico, que no ha sido posible sustituir completamente, con la evolución de la estadística y el apareamiento de nuevas y mejores herramientas de análisis, así como del equipo de cómputo, es posible construir modelos más complejos y un análisis más profundo de la información producto de las muestras obtenidas,

Anderson, Sweeney, Williams, Camm, y Cochran (2016) indican que un modelo de regresión describe la relación entre una variable dependiente y una o más variables independientes y agrega un término para el error. Los principales problemas en la construcción de modelos consisten en determinar la forma funcional adecuada de la relación y las variables independientes que se incluirán en el modelo, el proceso de producción de azúcar ya tiene identificadas las variables involucradas en su proceso, una revisión exhaustiva de las mismas podría determinar un tratamiento especial que puedan requerir dada la naturaleza de los mismos datos.

Asimismo, Authomathika (2021), indica que la calibración de los sensores de BX (Brix) se hace con una regresión lineal, los datos de laboratorio y de los medidores instalados en el tacho, una medición de Brix adecuada. En la producción de granos de azúcar con pureza adecuada es importante considerar que, aunque los sensores muestran datos con cierto nivel de significancia por

haber sido obtenidos a partir de una regresión, no resulta recomendable que se produzca una nueva regresión a partir de ellos.

Para Montanero (2008), el caso del modelo multivariante el cual no cuenta con una variable sino con n variables generadoras, sigue el mismo esquema que en el caso unidimensional, con la salvedad de que no existe una distribución que generalice unívocamente la distribución F . Para el análisis se puede usar la distribución matricial, la distribución de Wishart, la distribución de Hotelling y la distribución de Wilks, Lawley-Hotelling, Roy y Pillay que pueden entenderse como cuatro formas diferentes de generalizar la distribución F en el caso multivariante, esa diversidad de pruebas permite desarrollar la habilidad del analista de datos seleccionando la que mejor se adecue al modelo construido en base a su propia naturaleza.

Si la operación en los tachos no se lleva a cabo de forma adecuada en cualquiera de las etapas puede ocurrir una caramelización, el cambio de color de la masa, disolución de los cristales o distintos tamaños de granos de azúcar, lo que causaría pérdidas económicas.

Según Mulet-Hing y Fernández-Salazar (2016) la concentración que presenta templa, el nivel en la presión del vapor, el vacío producido por la inyección de agua y la medición realiza sobre la cantidad de masa son los controles principales en el proceso de cocimiento para lograr la cristalización del azúcar. De los parámetros descritos en los controles de la cristalización del azúcar se obtienen los datos que ayudan a parametrizar el proceso y mejorar la calidad del grano de azúcar.

Por su parte, Valle (2018) menciona que la aplicación de diferentes pruebas estadísticas a todos los datos y la correlación establecida permite dotar de

confiabilidad a los resultados de un análisis de regresión lineal, por ello en el estudio se deben desarrollar todas las pruebas estadísticas que sean necesarias para establecer los criterios que mejor se adapten a la solución de la problemática.

La evolución de las técnicas de producción de azúcar se ha mejorado debido a los procesos de automatización que utilizan herramientas de análisis estadístico.

2. MARCO TEÓRICO

2.1. Modelos de regresión

Cuando se tienen datos de los parámetros de un proceso de una población o muestra se puede realizar un análisis de regresión con el objetivo de construir un modelo que muestre la asociación entre las variables. La variable por predecir es la variable dependiente y la variable o variables predictoras son variables independientes. Cuando intervienen una variable a predecir y una predictora se le conoce al modelo regresión lineal simple y en el que hay una variable a predecir y dos o más predictoras se le conoce como modelo de regresión múltiple (Anderson *et al.*, 2016).

Para crear el modelo que represente el fenómeno a investigar se deben realizar procedimientos estadísticos entre ellos el método de mínimos cuadrados, con la ecuación obtenida se pueden realizar pronósticos, pero antes de ello se debe validar la confiabilidad y precisión, analizando los coeficientes de correlación y determinación los cuales proporcionan el nivel de ajuste de la distribución del modelo con respecto a la distribución teórica.

A continuación, se detallan las herramientas y procedimientos estadísticos necesarios para la predicción de un modelo de regresión que se ajuste con la mejora precisión al fenómeno investigado.

2.1.1. Muestreo

“El objetivo de cualquier encuesta por muestreo es realizar inferencias acerca de una población de interés, partiendo de la información obtenida de una muestra de dicha población” (Scheaffer, Mendenhall, y Ott, 1987, p. 15). Los estimadores puntuales encontrados con la información de a muestra sirven para estimar los valores de los parámetros poblacionales con determinado nivel de significancia.

“Un estimador es una función de variables aleatorias observables y quizás otras constantes conocidas, usadas para estimar un parámetro” (Scheaffer *et al.*, 1987, p. 15). Cuando el valor esperado del estimador es igual al del parámetro se dice que es insesgado, si se cuentan con varios estimadores del mismo parámetro se toma el de la varianza mas pequeña.

Se dice que una muestra es aleatoria cuando al seleccionar las n unidades muestrales todas tienen la misma probabilidad de ser seleccionada, esto es el diseño de muestreo irrestricto.

“Un diseño de encuesta por muestreo que es ampliamente usado debido principalmente a que simplifica el proceso de selección de la muestra se denomina muestreo sistemático” (Scheaffer *et al.*, 1987, p. 168). En el muestreo sistemático se selecciona un elemento de una lista en un intervalo apropiado y se sigue con los siguientes elementos en el mismo intervalo.

Este muestreo se utiliza para control de calidad industrial, el investigador selecciona una muestra sistemática 1-en-3, una de 1-en-5 o 1-en- k . Después de determinar un número aleatorio R entre 1 y K el cual establece que la muestra

este formada por las unidades numeradas $R, R+k, R+2k, \dots, R+(n-1)k$ (Lohr, 2000).

Rodríguez, Machado, y Villamarin, (2019) incian que para econtrar la muestra que es la cantidad de unidades que se analizarán para medir la característica de la calidad de la variable, se debe definir el tamaño de la submuestra y el número submustras o frecuencia. El tamaño de la submuestra puesde estar entre 15 y 25 datos si las variaciones son pequeñas en el control de la media aritmética.

Para determinar el número de submuestras se debe estimar un tamaño mínimo en un día de producción, se establece un nivel de confianza $(1-\alpha)$ para la media poblacional $\mu_{\bar{y}}$ de la variable analizada con el cual se optiene el estadístico z al verificar que los datos se distibuyen de forma normal, un valor frecuente para el nivel de confianza es del 95 %. Tambien se tiene que tomar en cuenta el margen de error máximo permitido E para la media estimada definido por $z_{\alpha/2}(\sigma \div \sqrt{n})$.

Anderson *et al.* (2016) determina la cantidad de elementos de la muestra utilizando un nivel de significancia expresado en el estadístico z , la varianza poblacional y un margen de error E , con la siguiente ecuación:

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} \quad (\text{Ec. 1})$$

“En procesos muy variables o en los que se cuenta con poca o ninguna información acerca de la dispersión del mismo, es recomendable aplicar un muestreo piloto, con el cual se determina su estimador insesgado s^2 ” (Rodríguez *et al.*, 2019, p. 5). Este muestreo piloto se realiza con una muestra preliminar .

2.1.2. Análisis de regresión lineal

Para Anderson *et al.* (2016) la ecuación que describe como se relaciona la variable x y y en la regresión lineal simple se representa con la siguiente ecuación:

$$y = \beta_0 + \beta_1 x + \epsilon \quad (\text{Ec. 02})$$

Donde β_0 es la intersección de la recta con el eje y , y β_1 es la pendiente del modelo de regresión, la cual puede ser positiva que indica una relación directamente proporcional entre x y y ; si es negativa indica una relación inversamente proporcional entre las variables; si $\beta_1 = 0$ no existe relación. La variable aleatoria ϵ representa la variabilidad de y que no puede explicarse en la ecuación.

Como no podemos obtener el valor de ϵ para el modelo se encuentra un valor esperado o la media de y , $E(y)$, con lo que se obtiene la siguiente ecuación:

$$E(y) = \beta_0 + \beta_1 x \quad (\text{Ec. 03})$$

Para Anderson *et al.* (2016) esta ecuación es válida únicamente cuando se cuenta con los valores poblacionales, para estimar los parámetros de β_0 y β_1 a partir de una muestra se deben calcular los estadísticos muestrales b_0 y b_1 respectivamente.

La siguiente ecuación representa la regresión estimada con datos muestrales:

$$\hat{y} = b_0 + b_1 x \quad (\text{Ec. 04})$$

Con la ecuación de la regresión estimada se puede obtener una predicción de y dado algún valor de x .

2.1.2.1. Método de mínimos cuadrados

Con un diagrama de dispersión se puede apreciar la asociación entre dos variables cuantitativas, es construido con sus pares ordenados (x,y) , los cuales pueden marcar una tendencia. Si esta es una recta, se podrá hacer una regresión lineal para representar la relación. Para Navidi (2006) la recta que ajusta los datos se conoce como recta de mínimos cuadrados.

“El método de mínimos cuadrados es un procedimiento en el que se usan los datos de la muestra para encontrar la ecuación de regresión estimada” (Anderson *et al.*, 2016, p. 603), esto se logra cuando se minimizan las diferencias entre cada valor de la muestra y_i y el estimado de la regresión \hat{y}_i . La ecuación que representa el método de mínimos cuadrados es:

$$\min \sum (y_i - \hat{y}_i)^2 \quad (\text{Ec. 05})$$

Para minimizar esta ecuación se debe de realizar un procedimiento de derivadas parciales respecto a b_0 y b_1 y se igualan a cero obteniendo las siguientes ecuaciones que encuentran los estimadores respectivos:

$$b_0 = \bar{y} - b_1 \bar{x} \quad (\text{Ec. 06})$$

$$b_1 = \left(\sum (x_i - \bar{x}) (y_i - \bar{y}) \right) \div \left(\sum (x_i - \bar{x})^2 \right) \quad (\text{Ec. 07})$$

Donde x_i es la variable independiente y y_i es la dependiente de la observación i -ésima respectivamente, \bar{x} y \bar{y} son las medias de las variables.

2.1.2.2. Coeficiente de determinación

El coeficiente de determinación es una cantidad que mide como un modelo explica la relación de una agrupación de datos, este es un estadístico de bondad de ajuste (Navidi, 2006).

“El coeficiente de determinación, r^2 se puede interpretar como el porcentaje de la suma total de cuadrados que se explica mediante el uso de la ecuación de regresión estimada” (Anderson *et al.*, 2016, p. 618).

La suma de cuadrados debido al error SCE es la cantidad que se minimiza con el método de mínimos cuadrados se encuentra con la siguiente ecuación:

$$SCE = \sum(y_i - \hat{y}_i)^2 \quad (\text{Ec. 08})$$

La STC es la suma total de cuadrados y se puede encontrar con la ecuación:

$$STC = \sum(y_i - \bar{y})^2 \quad (\text{Ec. 09})$$

La suma de cuadrados debido a la regresión SCR, se usa para medir cuánto se desvían de la media los valores estimados en la recta de regresión y se puede encontrar al relacionarlos como se muestra a continuación:

$$SCR = STC - SCE = \sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2 \quad (\text{Ec. 10})$$

La razón en entre la SCR y la STC es el coeficiente de determinación, la ecuación se expresa de la siguiente forma:

$$r^2 = \frac{SCR}{STC} \quad (\text{Ec. 11})$$

Este coeficiente se expresa en porcentaje y puede tomar valores entre 0 % y 100 % donde los valores cercanos al 100 % son un buen ajuste del modelo de regresión encontrado.

2.1.2.3. Coeficiente de correlación

Navidi (2006) define el coeficiente de correlación r como la asociación lineal entre dos variables. Para calcular este coeficiente, primero se encuentran las medias de la variable predictora x y la variable a predecir y , con estos estimadores se operan en la ecuación:

$$r = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\left(\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \right)} \quad (\text{Ec. 12})$$

Donde x_i es la variable independiente y y_i es la dependiente de la observación i -ésima respectivamente, \bar{x} y \bar{y} son las medias de las variables.

Para Navidi (2006) los valores de r varían entre -1 y 1, los valores distintos de cero y muy cercanos a los extremos indican mayor fuerza de relación y los valores cercanos a cero o el mismo cero indican que no hay correlación lineal.

El coeficiente de correlación se puede encontrar con el signo de la pendiente del modelo lineal y la raíz cuadrada del coeficiente de determinación (Anderson *et al.*, 2016).

$$r_{xy} = (\text{signo de } b_1) \sqrt{\text{Coeficiente de determinación}} \quad (\text{Ec. 13})$$

Cuando se cuentan con pocos datos y estos contienen algún atípico el resultado del coeficiente de correlación no es concluyente al modelo de regresión lineal.

Otro aspecto para tomar en cuenta es la confusión. “La confusión ocurre cuando hay una tercera variable correlacionada con las dos variables de interés lo que da como resultado una correlación entre ambas” (Navidi, 2006, p. 484), antes de que se concluya que las dos variables tienen una relación casual se debe descartar la posibilidad de confusión.

2.1.2.4. Supuestos del modelo

La ecuación del modelo de mínimos cuadrados es $y = \beta_0 + \beta_1 x + \epsilon$, este modelo es significativo para la asociación de las variables si se comprueban los supuestos del término del error ϵ los cuales se detallan a continuación:

El término del error ϵ es una variable aleatoria cuya medida, o valor esperado, es cero; es decir, $E(\epsilon) = 0$. Implicación. β_0 y β_1 son constantes, por lo tanto, $E(\beta_0) = \beta_0$ y $E(\beta_1) = \beta_1$; así, para un valor dado de x el valor esperado de y es $E(y) = \beta_0 + \beta_1 x$.

La varianza de ϵ , que se denota con σ^2 , es la misma para todo valor de x . Implicación. La varianza de y respecto a la recta de regresión y es igual a σ^2 y es la misma para todos los valores de x . (Anderson *et al.*, 2016, p. 622)

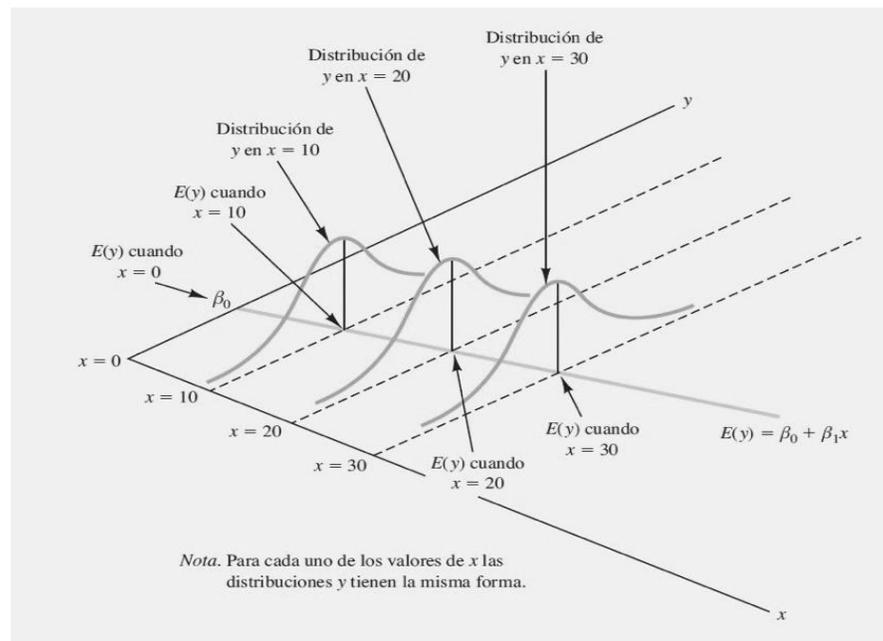
Esto significa que si la varianza no tiene un cambio estadísticamente significativo se está cumpliendo con la homocedasticidad.

Los valores del error ϵ son independientes. Implicación. El valor de ϵ correspondiente a un determinado valor de x no se relaciona con el valor de ϵ para cualquier otro valor de x ; por lo tanto, el valor de y correspondiente a un valor particular de x no se relaciona con el valor de y de ningún otro valor de x . (Anderson *et al.*, 2016, p. 622)

“El término error ϵ es una variable aleatoria distribuida normalmente. Implicación. Como y es una función lineal de ϵ , también y es una variable aleatoria con distribuida normalmente” (Anderson *et al.*, 2016, p. 622).

Estos supuestos y sus implicaciones se pueden apreciar gráficamente en la siguiente figura:

Figura 1. **Supuestos acerca del error ϵ**



Fuente: Anderson *et al.* (2016). *Estadística para negocios y economía*.

2.1.2.5. Prueba de significancia

“Para probar si existe una relación significativa de regresión, se debe realizar una prueba de hipótesis a efecto de determina si el valor de β_1 es distinto de cero” (Anderson, *et al.*, 2016, p. 623).

Al rechazar la hipótesis nula se concluye que las variables dependientes e independientes están relacionadas. La prueba de significancia usa las Prueba t y la Prueba F las cuales requieren una estimación de la varianza del error ϵ del modelo de regresión.

Para estimar el error cuadrático medio σ^2 se utiliza la siguiente ecuación:

$$ECM = s^2 = \frac{SCE}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2} = \frac{\sum(y_i - (b_0 + b_1 x_i))^2}{n-2} \quad (\text{Ec. 14})$$

Donde: SCE es la suma de los cuadrados de los residuales o errores y El error estándar de estimación s es la raíz cuadrada de ECM (Anderson, *et al.*, 2016).

- **Prueba t**

Con esta prueba se espera determinar que β_1 es distinto de cero, la prueba de hipótesis se plantea con $H_0: \beta_1 = 0$ y $H_a: \beta_1 \neq 0$. Cuando el valor p es menor que el nivel de significancia se rechaza H_0 , se podrá concluir que hay una relación significativa entre x y y (Anderson *et al.*, 2016). Esta prueba es posible únicamente cuando el modelo es lineal.

“Bajo los supuestos 1 al 4, las cantidades $\frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}}$ y $\frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$ tienen una distribución t de Student con n-2 grados de libertad” (Navidi, 2006, p. 512).

El estadístico de prueba para β_1 se define como:

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} \quad (\text{Ec. 15})$$

Donde: $\hat{\beta}_1$ que es un estimador insesgado de β_1 y $s_{\hat{\beta}_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$ es la estimación del error estándar de $\hat{\beta}_1$ con n-2 grados de libertad.

Navidi (2006) indica que “Los intervalos con un nivel de confianza de $100(1-\alpha)\%$ para β_0 y β_1 están dados por $\hat{\beta}_0 \pm t_{n-2, \alpha/2} \cdot s_{\hat{\beta}_0}$ $\hat{\beta}_1 \pm t_{n-2, \alpha/2} \cdot s_{\hat{\beta}_1}$ donde $s_{\hat{\beta}_0} = s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ $s_{\hat{\beta}_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$ ” (p. 513), y con esto se puede establecer los límites inferiores y superiores para el intercepto y la pendiente del modelo.

- **Prueba F**

Se desarrolla con la distribución F de Fisher siendo el estadístico calculado:

$$F = \frac{CMR}{ECM} \quad (\text{Ec. 16})$$

Donde:

- CMR = cuadrado medio de la regresión.
- ECM = error cuadrado medio.

Para valores grandes de F se rechaza H_0 ($\text{Valor } p \leq \alpha$), la correspondencia de la variable dependiente e independiente será significativa. Se contrasta el F calculado con el estadístico de prueba que depende del nivel de significancia que se esté utilizando.

Un análisis de varianza ANOVA resume los cálculos de la prueba con la información descrita en una tabla como la que se presenta a continuación.

Tabla II. **ANOVA para la prueba F**

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F	Valor p
Regresión	SCR	1	$CMR = \frac{SCR}{1}$	$F = \frac{CMR}{ECM}$	Rechazar H_0
Error	SCE	n-2	$ECM = \frac{SCE}{n-2}$		$Valor - p \leq \alpha$
Total	STC	n-1			

Fuente: Anderson *et al.* (2016). *Estadística para negocios y economía*.

○ **Prueba de normalidad**

Para Siegel y Castellan (1995) la prueba de Kolmogorov Smirnov mide el nivel de concordancia de la distribución observada y alguna distribución teórica específica como la distribución normal.

El estadístico de prueba para la distribución de probabilidad está dado por:

$$D = \max|F_o - F_e| \quad (\text{Ec. 17})$$

Donde:

- F_o es la frecuencia acumulada observada.
- F_e es la frecuencia acumulada calculada.

Siegel y Castellan (1995) indican los pasos para aplicar la prueba incluyen especificar la distribución acumulada teórica, organizar las F_o en una distribución acumulada relativa, encontrar el valor de D , verificar en la tabla de distribución el valor de D encontrado, si es menor que alfa, se rechaza H_o .

A continuación, se resumen los pasos de la prueba de Kolmogorov Smirnov:

- $H_o: F_o(x) = F_e(x)$.
- $H_a: F_o(x) \neq F_e(x)$ para al menos una x .
- $\alpha=1 \%, 5 \%, 10 \%$.
- $D_{\alpha/2}$ con tamaño de muestra n .
- Estadístico de prueba
- $D = \max|F_o - F_e|$
- Si $D > D_{\alpha/2}$ se rechaza H_o

2.1.3. Regresión múltiple

Cuando se necesitan tomar varios factores en consideración y se tienen p variables predictoras se utiliza la regresión múltiple. El modelo de regresión múltiple describe como se relaciona la variable dependiente con las variables independientes x_1, x_1, \dots, x_p y tiene la siguiente estructura:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (\text{Ec. 18})$$

Al igual que el modelo de regresión lineal ϵ no puede ser explicado en el modelo por lo que representa el error y los coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son los parámetros el valor esperado de y es:

$$E(y) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \text{ (Ec. 19)}$$

Anderson *et al.* (2006) establece que se usan los datos de la muestra aleatoria simple para encuentran los estimadores puntuales de los parámetros poblacionales, los cuales generan la ecuación de regresión múltiple estimada:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \text{ (Ec. 20)}$$

Para Navidi (2006) hay situaciones en las que una variable regresora no determina la solución a una problemática y en la práctica hay casos especiales de modelo de regresión múltiple en el cual se agregan términos con potencias de una sola variable, este es el modelo de regresión polinomial de la forma $y_i = \beta_0 + \beta_1x_i + \beta_2x_i^2 + \dots + \beta_px_i^p + \epsilon_i$. Estos casos especiales se deben de tomar en cuenta en este tipo de análisis.

2.1.3.1. Método de mínimos cuadrados

De forma similar a la regresión lineal, se utiliza el método de mínimos cuadrados, para obtener una ecuación con términos formados por distintas variables regresoras. Se requieren de datos de una muestra para encontrar los estimadores puntuales $b_0, b_1, b_2, \dots, b_p$ para minimizar el total de los cuadrados. Este proceso se basa en la solución de sistemas de ecuaciones donde se utiliza algebra matricial que viene implícito en algún software estadístico.

Según Navidi (2006), los estimadores más usados similares a los de la regresión lineal son: s^2 que es la varianza ϵ , el coeficiente de determinación R^2 y F .

El estimador de la varianza de ϵ se encuentra como:

$$s^2 = \frac{SCE}{n-p-1} \quad (\text{Ec. 21})$$

Con $SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. En donde n es el número de observaciones y p es el número de términos del modelo (Navidi, 2006).

Los cuatro supuestos del modelo que aplican de la misma manera que en regresión lineal: los errores $\epsilon_1, \dots, \epsilon_n$ son aleatorios e independientes, tienen una media de cero, todos tienen la misma varianza y están distribuidos normalmente.

Al cumplirse los supuestos el valor $\frac{b_p - \beta_p}{s_{b_p}}$ tienen una distribución t de Student con $n - p - 1$ grados de libertad, esta cantidad sirve para encontrar los intervalos de confianza y establecer las pruebas de hipótesis.

“El valor R^2 se calcula del mismo modo que r^2 en la regresión lineal simple” (Navidi, 2006, p. 560). La ecuación se escribe de la siguiente manera:

$$R^2 = \frac{\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = \frac{STC - SCE}{STC} = \frac{SCR}{STC} \quad (\text{Ec. 22})$$

“La hipótesis nula análoga es $H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$. [...]. Establece que ninguna variable independiente tiene alguna relación lineal con la variable dependiente” (Navidi, 2006, p. 560). Lo que se espera siempre es rechazar H_0 y se usa el estadístico F como se describe en la siguiente ecuación:

$$F = \frac{(STC - SCE) \div p}{STC \div (n - p - 1)} = \frac{SCR \div p}{STC \div (n - p - 1)} = \frac{SCR \div p}{s^2} \quad (\text{Ec. 23})$$

Donde: p y $n - p - 1$ son los grados de libertad en la distribución F de Fisher.

2.1.3.2. Comprobación de supuestos en la regresión múltiple

Se deben de comprobar la validez de los supuestos para los errores en el modelo de regresión múltiple, estos son los mismos que se deben cumplir en la regresión lineal simple, para Navidi (2006) “Estos diagnósticos son las gráficas de residuos contra valores ajustados, las de probabilidad normal de residuos y las de residuos contra el orden en que se hacen las observaciones” (p.564).

Un método gráfico da una pauta para identificar si se cumple el supuesto, se deben plantear pruebas de hipótesis para comprobarlo.

2.1.3.3. Confusión y multicolinealidad

Para el diseño de experimentos con un análisis factorial otros autores han afirmado lo siguiente:

La confusión consiste en que al estimar el efecto de bloque también se está calculando la interacción o interacciones utilizadas en su construcción. Mientras los efectos utilizados en la generación de los bloques sean de interacciones triples en adelante, el efecto confundido se puede atribuir sin mayor trámite a los bloques. (Gutiérrez y De la Vara, 2008, p. 218)

“La estructura de alias describe el patrón de confusión que ocurre en un diseño. También se dice que los términos que se confunden forman una estructura de alias” (Minitab, 2020, párr. 1).

Para Anderson *et al.* (2016) la multicolinealidad expresa la correlación entre las variables independientes y que los valores del coeficiente de correlación muestral mayores que +0.7 y menores que -0.7 para dos variables independientes presentan problemas potenciales de multicolinealidad. El efecto de las variables dependientes o independientes se analiza de forma conjunta al no poder dividirlos.

“La multicolinealidad en regresión es una condición que ocurre cuando algunas variables predictoras incluidas en el modelo están correlacionadas con otras variables predictoras” (Minitab, 2020, párr. 8).

Montgomery, Peck, y Vining (2006) sugieren que se pueden hacer un diagnóstico de multicolinealidad al encontrar los factores de inflación de la varianza (VIF, de *Variance inflation factor*). Si el VIF es mayor de 10 indica problemas de multicolinealidad.

Para encontrar el VIF se utiliza la ecuación:

$$VIF_j = \frac{1}{1-R_j^2} \quad (\text{Ec. 24})$$

Donde R_j^2 representa el coeficiente de determinación múltiple.

2.2. Cristalización de azúcar

El desarrollo del proceso se lleva a cabo formando condiciones de vacío al tacho designado para la cristalización, se considera la pureza y la concentración °Brix de la meladura y la miel A la cual se calcula y se alimenta al tacho. Se reduce la cantidad de agua de la mezcla para alcanzar el punto de saturación que determina el medidor de Brix del aparato y la experiencia del operador.

Cuando se alcanza el nivel de saturación se introduce al tacho la semilla que es un precursor para iniciar la conformación de cristales, en el lapso de 2 a 4 minutos comienzan a aparecer los nuevos cristales de tamaño esperado. Luego se introduce agua caliente al tacho por un periodo de 30 a 35 minutos, para que el cristal de azúcar alcance consistencia y uniformidad, evitando variaciones en el brix de la masa que puedan disolver los cristales.

Por último, se procede a alimentar con miel A en la cantidad determinada por cálculos de acuerdo con las prioridades de la mezcla y se corana con miel B hasta llegar la templa a 1800 pies cúbicos. Esta cantidad de pies cúbicos sirve para tres templeas de masa tercera.

“Dentro del proceso general de producción de azúcar reviste particular importancia la cristalización en los tachos, considerando las elevadas concentraciones y pureza de los materiales que se manejan en los mismos” (Mulet-Hing y Fernández-Salazar, 2016, párr. 5).

El proceso de la cocción de las mieles para generar el azúcar es uno de los puntos que requieren mayor atención ya que el producto final es un grano de azúcar que cumpla con el tamaño y consistencia que exigen los estándares de calidad requerida. En esta etapa se forman los cristales de azúcar.

Según Domínguez (2013) el termino pureza es utilizado como un indicador del contenido de azúcar existente en la solución. La concentración de sacarosa con la que se mide la pureza se incrementa con la evaporación, obtener calidad en el producto terminado al evaporar varias veces la solución que alimenta el tacho representa un costo que afecta la producción.

En los tachos de cocimiento se produce la cristalización del azúcar a alta temperatura. Para trabajar a temperaturas menores a los 100 °C y facilitar que se formen los cristales, se provoca vacío por presión al tacho (Golato, 2014).

El cocimiento se realiza por etapas llamadas templeas, que no es más que la masa con diferentes concentraciones sacarosa, se pueden realizar de una a cuatro etapas, siendo la más común la de tres. (Golato, 2014).

La primera masa cocida del melado se le denomina masa cocida A y la miel separada en la etapa de centrifugado se le llama miel A, se utiliza para hacer crecer la masa cocidas de otras templeas. Cuando se alcanza la templea por segunda vez se obtiene la masa B y la miel B que extrae en las centrifugadoras. La última miel que se produce en las templeas es la melaza.

La sobresaturación se da cuando se supera el nivel de concentración normal de sacarosa en la concentración. Los tachos operan en una zona sobresaturada entre 1.25 y 1.40, punto ideal para la formación de cristales. En un tacho se desarrollan cuatro pasos: concentración, semillado para cristalizar, crecimiento del grano y apriete de la masa cocida (Golato, 2014).

El desarrollo de la masa de tercera tiene por objetivo alimentar la masa de tercera con miel de segunda para desarrollar el grano formado lo cristales hasta el tamaño adecuado que permita el mayor agotamiento de miel. Las masas de tercera (de la tercera templea) realizan el trabajo de agotamiento de miel B y crecimiento inicial de los cristales formados por el semillamiento para permitir la formación de magma C y la salida de miel fina durante todo el periodo de zafra. Esta tarea es realizada por el operador de tachos de tercera (Corea, 2008).

3. PRESENTACIÓN DE RESULTADOS

3.1. Identificación de las variables

De acuerdo con los objetivos propuestos se presentan los siguientes resultados:

- Objetivo 1: identificar las variables que influyen en el proceso de cristalización del azúcar, utilizando metodología de Análisis de varianza

Para validar las variables se calculó el tamaño n de la muestra con un nivel de confianza $(1-\alpha)$ del 95 % y con una desviación estándar $s=5.85$ proveniente de datos de estudios históricos al no contar con una desviación estándar poblacional del proceso. El margen de error que se acepta es del 2 °Brix. Al ingresar la información en la ecuación 1 se obtiene:

$$n = \frac{(Z_{\alpha/2})^2 \sigma^2}{E^2} = \frac{(1.96)^2 (5.85)^2}{2^2} = 32.86 = 33$$

Para la calibración del dispositivo de medición, el Departamento de Automatización del ingenio azucarero tiene un protocolo establecido con muestras de 24 datos del tacho de refinería para calibrar el equipo, por lo que no fue posible obtener los 33 datos sugeridos por la ecuación. Adicionalmente se cuentan con datos tomados para rectificar algunos valores fuera de normal con pruebas hechas en laboratorio.

Se realizó un análisis de los datos históricos con una tabla combinada de Excel para 190 datos con los cuales se verificó la asociación de la variable de respuesta °Brix y las predictoras: fase, fase compensada, temperatura y atenuación con la etapa de la templa.

Tabla III. **Promedio de las variables en cada etapa de la templa**

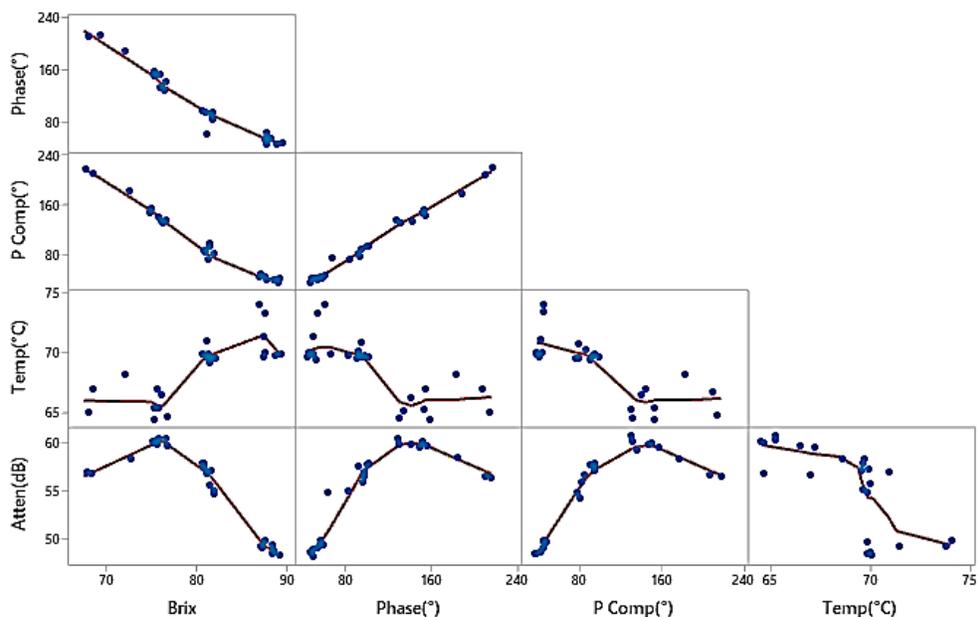
Etapa de Templa	Promedio de Brix	Promedio de P (°)	Promedio de P Comp (°)	Promedio de T(°C)	Promedio de At(dB)
Carga	75.4	220.3	222.2	64.9	58.7
Concentración	77.4	169.9	168.2	65.5	55.6
Semillamiento	81.2	139.3	131.1	68.7	52.1
Directa	86.3	125.0	113.7	68.8	50.4
Aprete previo	89.7	87.7	79.6	67.9	49.7
Aprete final	89.0	62.0	50.9	70.0	42.3
Descarga	91.5	73.2	60.8	69.6	45.2

Fuente: elaboración propia.

Los resultados mostraron que hay una relación inversa entre la variable de salida °Brix y la fase, la fase compuesta y la atenuación, cuando el Brix aumentas estas disminuyen, la relación de la variable de salida con la temperatura es directamente proporcional. Con esta información se descarta la etapa como una variable regresora ya que su función es categorizar a la variable de salida.

De forma visual se puede comprobar la relación al realizar diagramas de dispersión para cada relación. Para apoyar en los todos cálculos se utilizó el software Minitab®, R Studio e InfoStat.

Figura 2. Gráfica de matriz de todas las variables

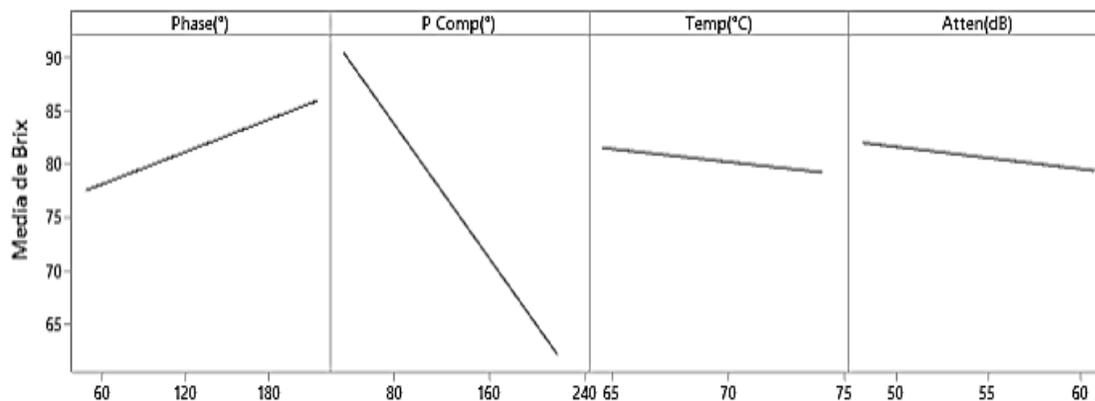


Fuente: elaboración propia.

La primera columna °Brix muestra la relación con cada variable, se puede ver que no tiene una relación aparente con la temperatura. En la segunda columna se observa que Phase guarda relación con P Comp y con Atten. En la tercera columna destaca que P Comp no tiene relación aparente con Temp. Esta gráfica es de mucha utilidad para buscar indicios de multicolinealidad entre las variables. La relación fase compensada y brix tiene tendencia cuadrática.

Después se procedió a realizar una gráfica de interacciones de las variables predictoras para confirmar qué variable influye en el modelo.

Figura 3. Efectos principales para la variable de salida Brix



Fuente: elaboración propia.

Las variables temperatura y atenuación no muestran un cambio significativo, se mantienen. La variable que más influye en el modelo es la fase compensada

Tabla IV. Análisis de varianza para tacho de refinería

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	4	916.809	229.202	381.16	0.000
Phase (°)	1	1.235	1.235	2.05	0.168
P Comp (°)	1	11.553	11.553	19.21	0.000
Temp (°C)	1	2.405	2.405	4.00	0.060
Atten (dB)	1	7.581	7.581	12.61	0.002
Error	19	11.425	0.601		
Total	23	928.234			

Fuente: elaboración propia.

Según los resultados de ANOVA se confirma que la variable regresora fase compensada es una de la que más aporta al modelo de regresión por lo que tiene mayor correlación con la repuesta °Brix.

La señal del dispositivo de medición se ve afectado por los cambios de temperatura del medio. Como resultado, el cambio de fase tiene que compensarse por temperatura. Para ello se utiliza una ecuación para obtener la fase compensada la cual la define la referencia de temperatura T_{Ref} , como el coeficiente de temperatura T_{k_n} . Este resultado parece ser evidente en la no significancia de estas variables en el ANOVA.

$$f_c = f - T_{k_3}(T_m - T_{Ref})^2 - T_{k_2}(T_m - T_{Ref})^2 - T_{k_1}(T_m - T_{Ref})$$

En los ingenios azucareros T_{Ref} debe ajustarse a la temperatura del punto de siembra de 65 °C, y el coeficiente de temperatura $T_k = + 3.0$ °C. Este dato fue importante describirlo porque define la relación entre las variables regresoras. Esta información causa dudas sobre la dependencia de la variable fase compensada con las variables temperatura y fase por lo que se validó un modelo que solo incluya esta variable.

Para la construcción del modelo de mejor ajuste se procedió a verificar la interacción de las variables tomando los criterios C_p de Mallows, el AIC y el BIC así como la multicolinealidad a través del VIF y como criterio de bondad de ajuste el mejor R cuadrado ajustado.

Tabla V. **Coefficientes para la construcción del modelo de regresión**

Término	Coef.	EE del coef.	Valor T	Valor p	FIV
Constante	119.10	9.50	12.53	0.000	
Phase (°)	0.0501	0.0350	1.43	0.168	123.68
P Comp (°)	-0.1574	0.0359	-4.38	0.000	148.57
Temp(°C)	-0.243	0.122	-2.00	0.060	3.97
Atten(dB)	-0.2063	0.0581	-3.55	0.002	2.64

Fuente: elaboración propia.

Para cada coeficiente se realizó la prueba de hipótesis $H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$ para establecer que ninguna variable independiente tiene alguna relación lineal con la variable dependiente, esperando rechazar H_0 . Se determina que los coeficientes de la temperatura y fase tienen un valor p mayor que el nivel de significancia $\alpha=0.05$ por lo que se establece que no son significativas, pero como los coeficientes de la atenuación y fase compuesta son significativos no se rechaza H_0 .

Se realizó un diagnóstico de multicolinealidad al encontrar los factores de inflación de la varianza (FIV), se encontró en la tabla V que la fase y la fase compuesta tienen valores mayores a 10 lo que significa que hay problemas de multicolinealidad en el modelo que incluye todas las variables predictoras.

Para el modelo que incluye todas las variables se establece los datos presentados en la siguiente tabla:

Tabla VI. **Resumen del coeficiente de determinación R²**

S	R-cuad.	R-cuad. (ajustado)	R-cuad.(pred)
0.775458	98.77 %	98.51 %	97.30 %

Fuente: elaboración propia.

La tabla muestra el valor de coeficiente de determinación R cuadrado de 98.77 % y un coeficiente de determinación ajustado de 98.51 % el cual es un valor muy bueno, pero es un resultado engañoso al tener un FIV elevado. Se procedió a quitar del modelo la variable de la fase compuesta P Comp (°) porque se determinó que esta variable depende de la fase y la temperatura, por lo cual genera una multicolinealidad, esto generó un nuevo análisis de las variables.

Tabla VII. **Coeficientes sin la variable fase compuesta**

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	101.0	11.8	8.54	0.000	
Phase (°)	-0.10132	0.00737	-13.75	0.000	2.88
Temp (°C)	0.068	0.137	0.50	0.626	2.62
Atten (dB)	-0.2595	0.0785	-3.30	0.004	2.53

Fuente: elaboración propia.

Se determinó que este modelo ya no tiene problemas de multicolinealidad al tener valores de FIV menores a 10 y la bondad de ajuste de R-cuad (ajustado) de 97.15 % es muy buena. Sin embargo, se tiene que el coeficiente de Temp no es significativo lo cual lleva a la eliminación de este coeficiente del modelo.

Tabla VIII. **Coefficientes de fase y atenuación**

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	106.60	3.53	30.16	0.000	
Phase (°)	-0.10306	0.00636	-16.20	0.000	2.22
Atten(dB)	-0.2730	0.0723	-3.77	0.001	2.22

Fuente: elaboración propia.

En esta tabla se muestran los resultados de los coeficientes de la ecuación de regresión lineal múltiple que cumple con los criterios del FIV, las pruebas de hipótesis para los coeficientes.

Como se indicó sobre la ecuación que forma la fase compuesta que depende de la temperatura y de la fase se planteó un segundo modelo que incluya la fase compensada y la atenuación y se validó por medio de los contrastes, y se determina un resultado significativo, esta asociación cumple con la prueba de hipótesis y la multicolinealidad.

Tabla IX. **Coefficientes de fase compuesta y atenuación**

Término	Coef	EE del coef.	IC de 95 %	Valor T	Valor p	FIV
Constante	101.56	2.88	(95.58, 107.55)	35.29	0.000	
P Comp(°)	-0.10208	0.00480	(-0.11207, -0.09209)	-21.24	0.000	2.41
Atten(dB)	-0.1954	0.0583	(-0.3167, -0.0742)	-3.35	0.003	2.41

Fuente: elaboración propia.

3.2. Construcción del modelo

- Objetivo 2: construir el modelo que mejor se ajusta al proceso interno programado en el sistema automático por medio del análisis de regresión

Según el análisis de los contrastes y la prueba de multicolinealidad se establecen tres modelos en relación con las variables regresoras combinadas como se muestra en la siguiente tabla.

Tabla X. **Combinación de variables regresoras para cada modelo**

Núm.	Modelo	Regresoras	Respuesta
1	Modelo de primer orden con dos variables (multilineal)	Fase (Phase) y atenuación (Atten)	Brix
2	Modelo de primer orden con dos variables (multilineal)	Fase compensada (Phase Comp) y Atten	Brix
3	Modelo de segundo orden en una variable (cuadrático)	Fase compensada (Phase Comp)	Brix

Fuente: elaboración propia.

3.2.1. Modelo 1

Para construir el primer modelo de regresión lineal múltiple de mejor ajuste se tomó el método de mínimos cuadrados de los resultados del ANOVA mostrado en la siguiente tabla.

Tabla XI. **Análisis de varianza para el modelo de regresión**

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	904.97	97.49 %	904.97	452.487	408.51	0.000
Phase (°)	1	889.20	95.79 %	290.75	290.752	262.50	0.000
Atten(dB)	1	15.78	1.70 %	15.78	15.777	14.24	0.001
Error	1	23.26	2.51 %	23.26	1.108		
Total	3	928.23	100.00 %				

Fuente: elaboración propia.

Se realizó la prueba de hipótesis de contrastes y se establece que el valor p para la variable fase y atenuación es menor que 0.05, lo que indica que los coeficientes para estas variables son significativos en el modelo.

De la tabla VIII se obtiene los coeficientes que formaron la ecuación de regresión lineal múltiple la cual se presenta a continuación.

$$B_x = -0.10306f - 0.2730A_t + 106.60$$

El coeficiente de determinación R^2 para el modelo propuesto es de 97.49 % y este indica que la relación de agrupación de las variables es muy buena. Se debe confirmar la validez de los supuestos estadísticos del modelo, los cuales se analizan posteriormente.

3.2.2. Modelo 2

Para el segundo modelo se realizó el siguiente análisis de varianza:

Tabla XII. **Análisis de varianza para el modelo 2**

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	914.27	98.50 %	914.27	457.14	687.63	0.000
P Comp (°)	1	906.81	97.69 %	300.05	300.05	451.34	0.000
Atten(dB)	1	7.47	0.80 %	7.47	7.47	11.23	0.003
Error	21	13.96	1.50 %	13.96	0.66		
Total	23	928.23	100.00 %				

Fuente: elaboración propia.

Al realizar la prueba de hipótesis de contrastes, se establece que el valor p para la variable fase y atenuación es menor que 0.05, lo que indica que los coeficientes para estas variables son significativos en el modelo.

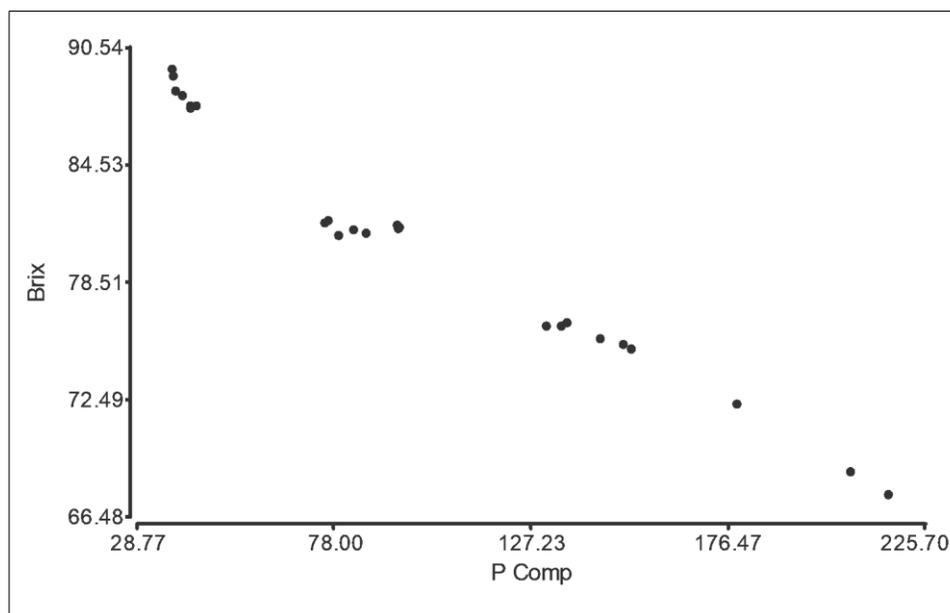
De la tabla IX se obtiene los coeficientes que formaron la ecuación de regresión lineal múltiple la cual se presenta a continuación.

$$B_x = -0.10208f_c - 0.1954A_t + 101.56$$

3.2.3. Modelo 3

Para el tercer modelo se toma en cuenta que el aporte de la atenuación para el modelo 2 es muy baja y la tendencia de los puntos en el diagrama de dispersión de los datos al relacionar la variable regresora fase compensada con la respuesta Brix. Se establece que hay una relación de orden mayor y se exploró una de segundo orden.

Figura 4. **Dispersión Brix contra fase compensada**



Fuente: elaboración propia.

Al revisar la información proporcionada para el tacho muestreado se establece que los parámetros internos del dispositivo de medición Promtec establecen una función de orden dos, por lo que se procedió a realizar el análisis para este modelo.

Tabla XIII. **Coefficientes para la fase compuesta del modelo cuadrático**

Término	Coef	EE del coef.	IC de 95 %	Valor T	Valor p	FIV
Constante	93.837	0.723	(92.333,95.340)	129.80	0.000	
P Comp	-0.1558	0.0143	(-0.1855,-0.1260)	-10.87	0.000	19.77
(PComp) ²	0.000178	0.000060	(0.000053,0.000303)	2.96	0.007	19.77

Fuente: elaboración propia.

Al validar los contrastes se confirma la significancia de la variable, se procedió a realizar el análisis de varianza que a continuación se detalla.

Tabla XIV. **Análisis de varianza del modelo de orden dos**

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	913.13	98.37 %	913.13	456.56	634.60	0.000
P Comp (°)	1	906.81	97.69 %	85.06	85.06	118.23	0.000
(P Comp) ²	1	6.32	0.68 %	6.32	6.32	8.78	0.007
Error	21	15.11	1.63 %	15.11	0.72		
Total	23	928.23	100.00 %				

Fuente: elaboración propia.

Al realizar la prueba de hipótesis de contrastes, se establece que el valor p para la variable fase y atenuación es menor que 0.05, lo que indica que los coeficientes para estas variables son significativos en el modelo.

De la tabla XIII se obtiene los coeficientes que formaron la ecuación de regresión lineal múltiple la cual se presenta a continuación.

$$B_x = 0.000178f_c^2 - 0.1558f_c + 93.837$$

3.3. Validación del modelo

- Objetivo 3: validar el modelo propuesto por medio de pruebas de hipótesis que verifiquen los supuestos de la ecuación obtenida.

3.3.1. Modelo 1

Para cada uno de los modelos se verificaron los supuestos de bondad de ajuste con el coeficiente de determinación, la normalidad de forma gráfica y confirmado de forma numérica, la independencia y la homocedasticidad.

El primer modelo incluye la fase y la atenuación descrito como:

$$B_x = -0.10306f - 0.2730A_t + 106.60$$

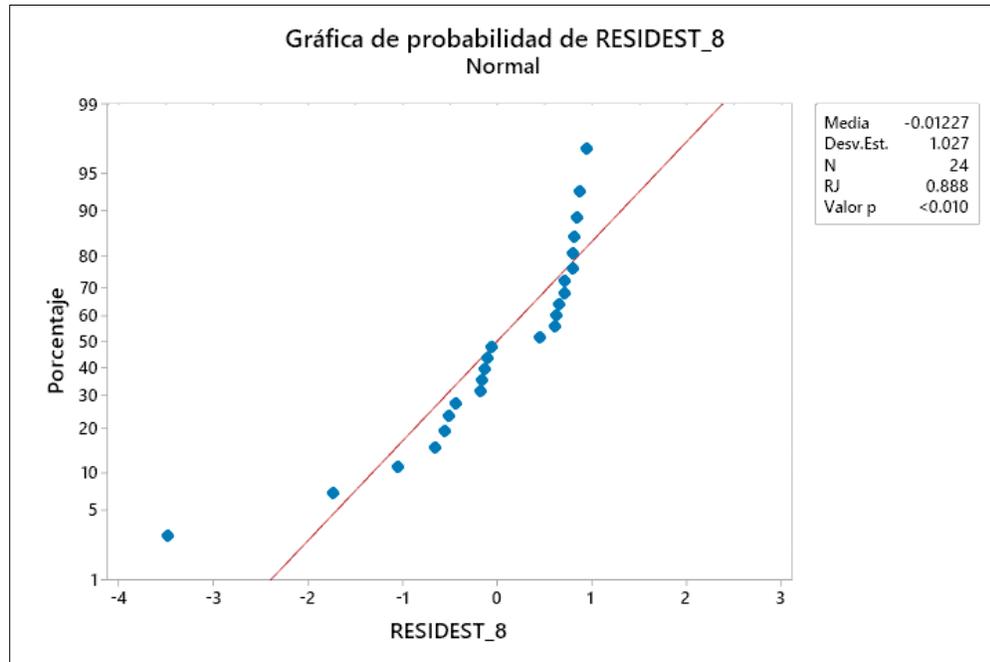
3.3.1.1. Bondad de ajuste

Lo primero que se verificó es la bondad de ajuste del modelo obtenido por medio del coeficiente de determinación R^2 ajustado de 97.26 % que significa que tienen mayor fuerza de relación por lo que el modelo se ajusta significativamente a los datos. El valor $F = 408.51$ indica que el modelo es significativo, según la prueba F , que para valores F comparado con $F_{\alpha=199}$, con un nivel de significancia de 5 %.

3.3.1.2. Prueba de normalidad

Se comprobó la validez de los supuestos del término del error. La primera prueba que se realizó es la de normalidad para ello primero se hizo un análisis gráfico de Q-Q *plot* de los residuos para verificar si todos los puntos siguen la recta que representa una distribución normal.

Figura 5. Gráfico de probabilidad normal del modelo 1



Fuente: elaboración propia.

A continuación, se aplicó la prueba de Kolmogorov Smirnov con los siguientes pasos:

- H_0 : los datos proceden de una distribución normal
- H_a : los datos no proceden de una distribución normal
- $\alpha=5\%$
- $D = 0.18635$; valor $p=0.3329$
- Si $Valor\ p < 0.05$ Se rechaza H_0

El valor p del test Kolmogorov Smirnov debe ser mayor que el nivel de significancia para que supere la prueba de normalidad, es decir los residuos tienen una distribución normal, se establece que el valor $p > 0.05$. Sin embargo,

este resultado es incorrecto ya que la prueba de Kolmogorov se aplica para muestras mayores de 50 porque asume que la media y la desviación estándar es poblacional.

Como los datos muestreados son menores de 50 se utiliza la prueba con la corrección de Lilliefors utilizando el software R *Rstudio* con la que se obtuvo el resultado mostrado en la figura.

Figura 6. **Test de Lilliefors con código de Rstudio**

```
> library(nortest)
warning message:
package 'nortest' was built under R version 4.0.3
> lillie.test(Prueba_de_KS$RESIDEST_8)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  Prueba_de_KS$RESIDEST_8
D = 0.18635, p-value = 0.03064
```

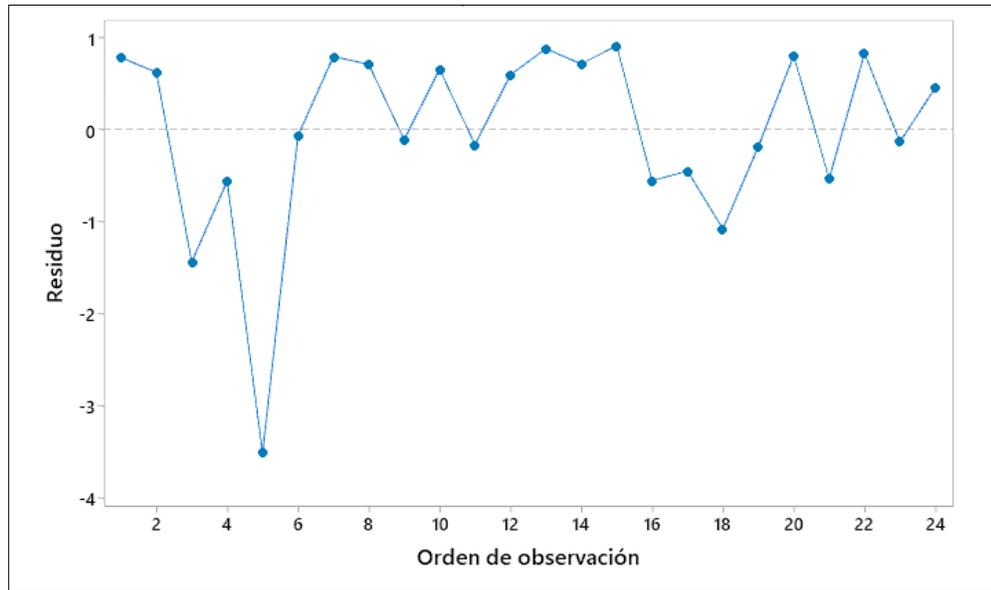
Fuente: elaboración propia.

Al realizar la comparación el valor p es menor que alfa, esto indica que los datos no proceden de una distribución normal.

3.3.1.3. Prueba de independencia

Para verificar la independencia del error aleatorio se realizó un gráfico de residuos y el orden de las observaciones, para establecer independencia en los residuos de los datos, al confirma la correlación entre sí. Si los datos están correlacionados no son independientes.

Figura 7. **Residuos contra orden del modelo 1**



Fuente: elaboración propia.

Al observar la gráfica se ve que tiene ciertas tendencias o patrones de correlación entre la observación 3, 4, 5 y 6 estos nos diría que están correlacionados y por lo tanto no son independientes.

Para verificar este supuesto se utilizó la prueba de autocorrelación de Durbin-Watson para evaluar si los errores de las observaciones adyacentes están correlacionados. Se utiliza el estadístico D, “Si $D > D_U$, no existe correlación; si $D < D_L$, existe una correlación positiva; si D se encuentra entre los dos límites, la prueba no es concluyente” (Minitab, 2019, párr. 3). Los valores de D_U y D_L se obtienen en la tabla de Savin y White que toma en cuenta la cantidad de términos del modelo y el número de muestras.

Tabla XV. **Prueba Durbin-Watson modelo 1**

Muestra	Términos	D_L	D_U	D	Resultado
24	3	1.18781	1,54639	1.64042	D > D_U

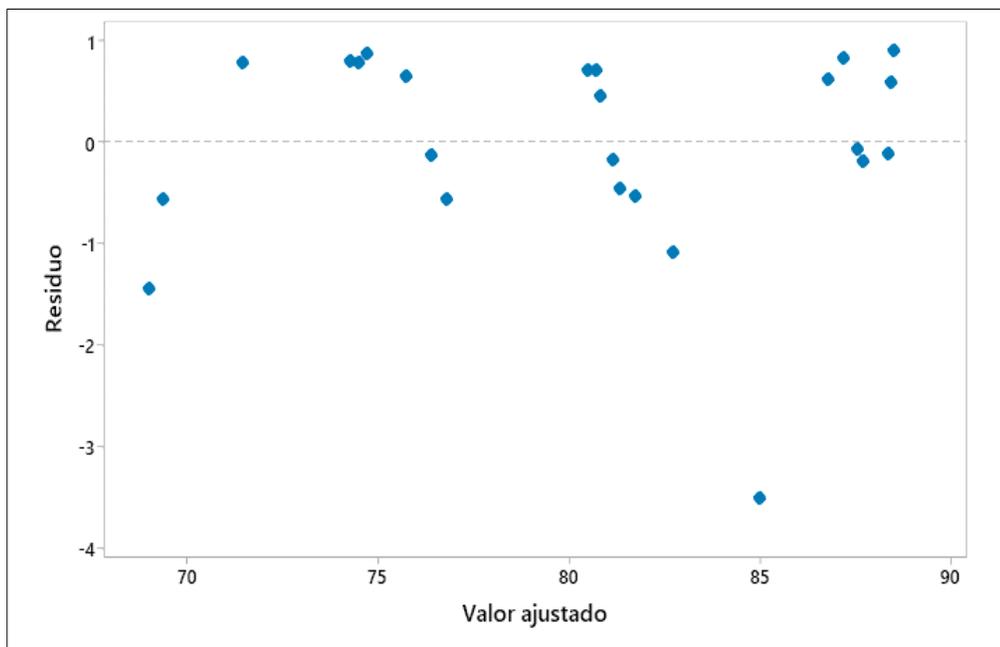
Fuente: elaboración propia.

El resultado de la prueba indica que no existe correlación en los errores de la observación, existe independencia en los residuos.

3.3.1.4. Prueba de homocedasticidad

La verificación de este supuesto trata de que los residuos o error tienen una varianza constante, lleva a revisar el gráfico de Residuos vs. Ajuste, en este gráfico se observa una nube de puntos disperso los cuales deben estar encerrados en una franja horizontal o tendencia tanto positiva como negativa, si estos presentan una tendencia se dice que son heterocedásticos y por lo tanto las varianzas no son iguales.

Figura 8. Residuos contra valor ajustado del modelo 1



Fuente: elaboración propia.

En el gráfico se observa que no hay un patrón definido, al trazar una recta horizontal en 1 y -1 quedan puntos fuera de estas rectas lo que significaría una posible variación, pero de forma general se puede decir los errores tienen homocedasticidad.

Para confirmar este resultado gráfico se utilizó la prueba de Breusch-Pagan que “consiste en ajustar un modelo de regresión lineal con variable DE respuesta dada por residuales del modelo original al cuadrado y como covariables las variables del modelo original” (Hernández y Mazo, 2020, párr. 2). El estadístico tiene una distribución chi cuadrado.

Se plantearon las siguientes hipótesis:

- H_0 : los errores tienen varianza constante.
- H_a : los errores no tienen varianza constante.

Utilizando el complemento XLSTAT en Excel se obtuvo el siguiente resultado

Tabla XVI. **Resultado de prueba Breush-Pagan modelo 1**

Estadístico	Valor
LM (Valor observado)	0.116
LM (Valor crítico)	5.991
GL	2
valor-p (bilateral)	0.944
Alfa	0.05

Fuente: elaboración propia.

Puesto que el valor-p=0.944 calculado es mayor que el nivel de significación $\alpha=0.05$, no se puede rechazar la hipótesis nula H_0 . Lo que indica que los errores tienen varianza constante.

3.3.2. Modelo 2

Este modelo lo constituye la fase compensada y la atenuación $B_x = -0.10208f_c - 0.1954A_t + 101.56$. Los resultados de las pruebas se detallan a continuación:

3.3.2.1. Bondad de ajuste

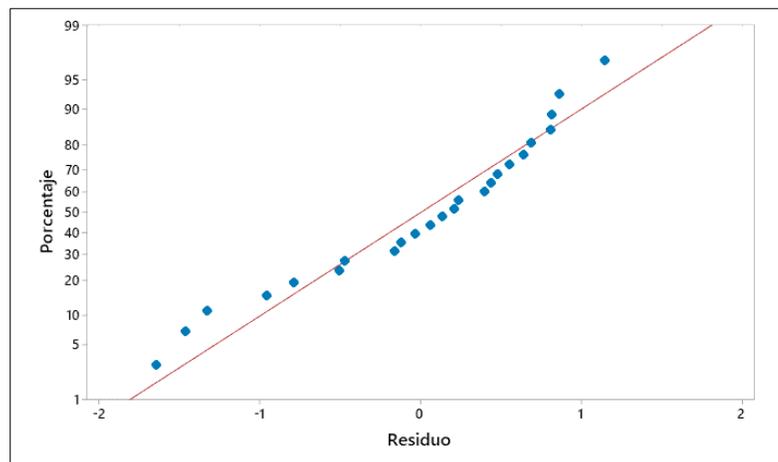
Se verificó es la bondad de ajuste del modelo obtenido por medio del coeficiente de determinación R^2 ajustado de 98.35 %, que significa que los

grados Brix se explican en un 98.35 % con el modelo que relaciona la fase compensada y la atenuación. El modelo explica muy bien la toda la variabilidad de los datos. El valor $F = 687.63$ indica que el modelo significativo, según la prueba F.

3.3.2.2. Prueba de normalidad

Se comprobó la validez de los supuestos del término del error. La primera prueba que se realizó es la de normalidad para ello primero se hizo un análisis gráfico de Q-Q *plot* de los residuos para verificar si todos los puntos siguen la recta de distribución normal.

Figura 9. **Gráfica de probabilidad normal del modelo 2**



Fuente: elaboración propia.

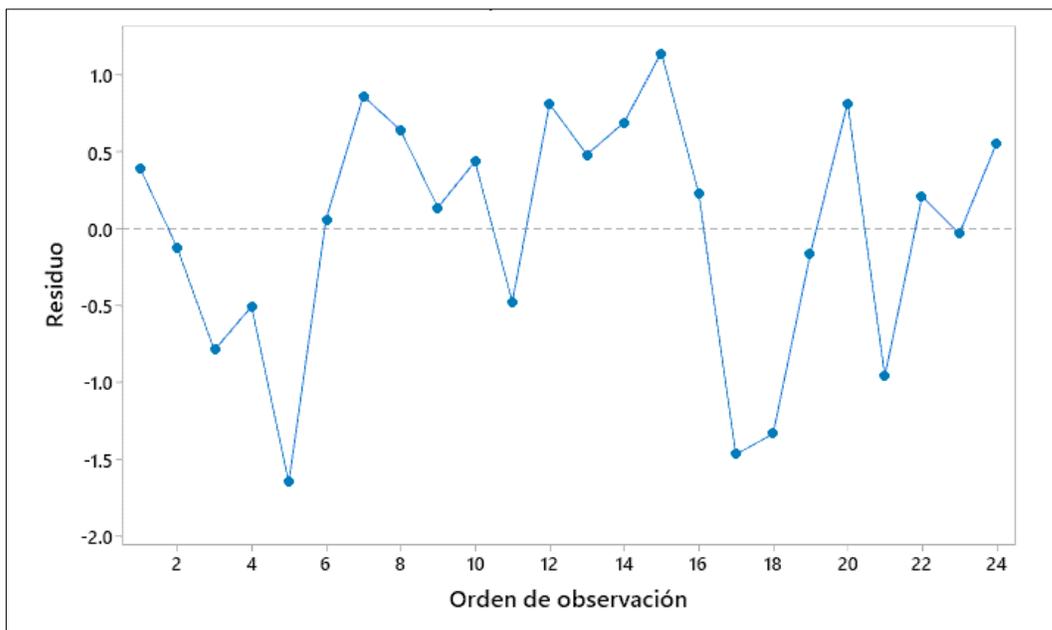
A continuación, se aplicó la prueba de Kolmogorov Smirnov con la corrección de Lilliefors con los pasos del modelo 1 para la hipótesis nula como H_0 : los datos proceden de una distribución normal. El estadístico $KS = 0.12934$, $p\text{-valor} = 0.3756$, al realizar la comparación el valor p es mayor que alfa:

$0.3756 > 0.05$, esto indica que los datos proceden de una distribución normal al no poder rechazar H_0 . Cumple con el supuesto de normalidad.

3.3.2.3. Prueba de independencia

Para verificar la independencia del error aleatorio se realizó el gráfico de residuos y el orden de las observaciones.

Figura 10. Residuos contra orden del modelo 2



Fuente: elaboración propia.

Al observar la gráfica se ve que tiene ciertas tendencias o patrones de correlación entre la observación 2, 3, 4 y 5 estos nos diría que están correlacionados y por lo tanto no son independientes. Se procede con la prueba de Durbin-Watson, como el valor de la muestra es el mismo que el modelo 1 al

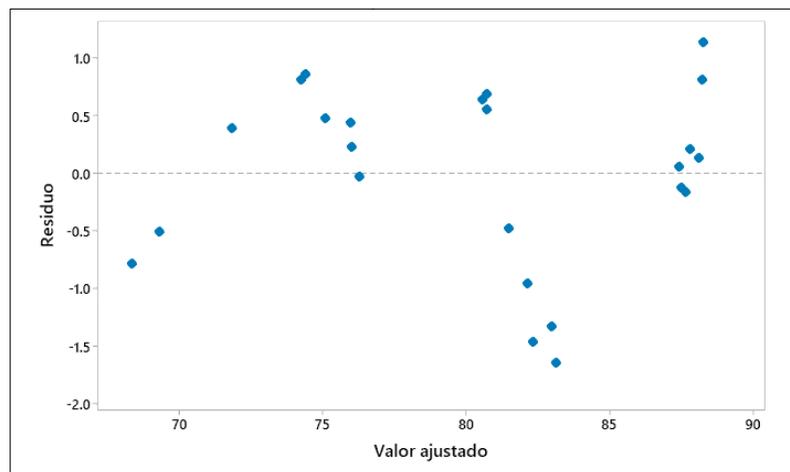
igual que los términos se compara el valor obtenido $D=1.4265$ de la tabla XIII dando como resultado que $D_L < D < D_U$.

El resultado de la prueba indica que no es concluyente por lo que no se puede determinar si existe o no correlación en los errores de la observación. Por consiguiente, no se pudo establecer que los datos son o no independientes.

3.3.2.4. Prueba de homocedasticidad

Al realizó gráfico de residuos contra ajuste y revisar la tendencia de los puntos se establece que los errores tienen homocedasticidad o que las varianzas son iguales.

Figura 11. Residuos contra valor ajustado modelo 2



Fuente: elaboración propia.

Se utilizó la prueba de Breush-Pagan con la hipótesis nula H_0 : los errores tienen varianza constante. El valor-p obtenido fue de 0.851 el cuál es mayor que

el nivel de significación $\alpha=0.05$, no se puede rechazar la hipótesis nula lo que indica que el error del modelo tiene varianza constante.

3.3.3. Modelo 3

Este modelo es una variante del modelo de regresión lineal múltiple y quedó formado por la fase compensada con un término lineal y otro cuadrático más el término constante.

$$B_x = 0.000178f_c^2 - 0.1558f_c + 93.837$$

A continuación, se presentan el resumen de las pruebas realizadas:

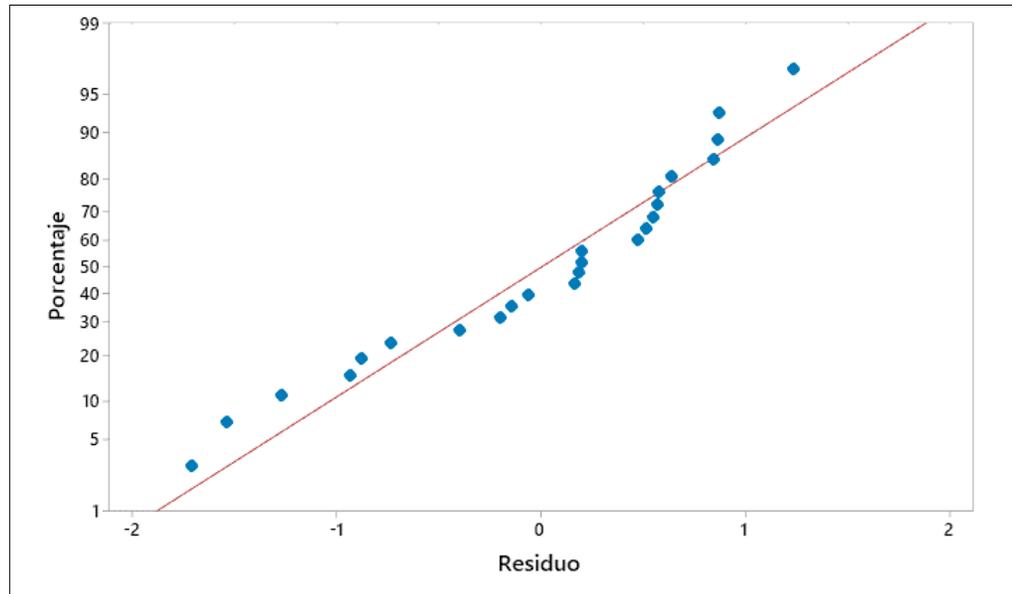
3.3.3.1. Bondad de ajuste

Se verificó es la bondad de ajuste del modelo obtenido por medio del coeficiente de determinación R^2 ajustado de 98.22 %, por lo que el modelo para obtener los grados brix se explica en un 98 % con la fase compensada. El valor $F = 634.60$ indica que el modelo significativo.

3.3.3.2. Prueba de normalidad

Se comprobó la validez de los supuestos del término del error. La primera prueba que se realizó es la de normalidad para ello primero se hizo un análisis gráfico de Q-Q *plot* de los residuos para verificar si todos los puntos siguen la recta de distribución normal.

Figura 12. **Gráfico de probabilidad normal para el modelo 1**



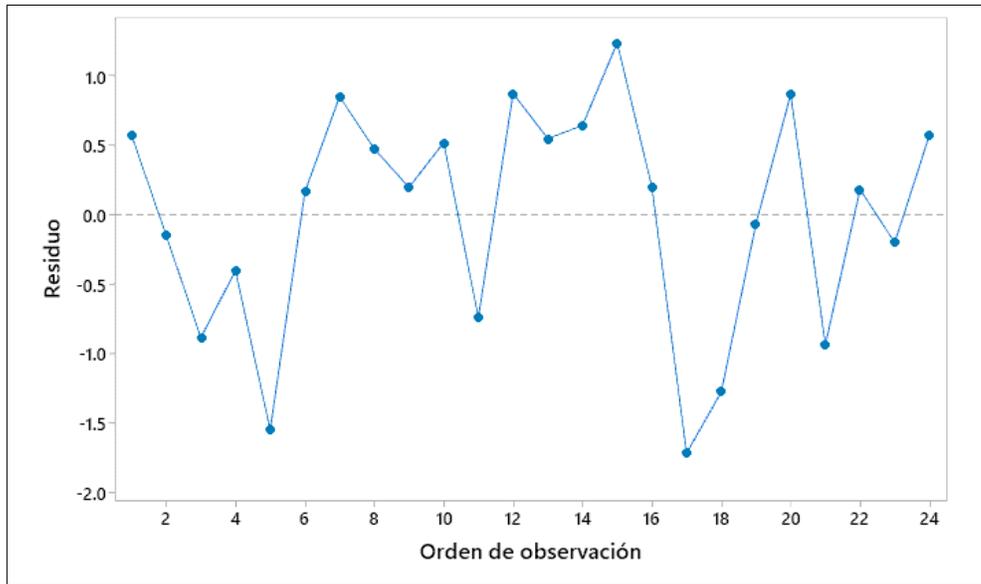
Fuente: elaboración propia.

A continuación, se aplicó la prueba de Kolmogorov Smirnov con la corrección de Lilliefors con los pasos del modelo 1 para la hipótesis nula como H_0 : los datos proceden de una distribución normal. El estadístico $KS = 0.16703$, $p\text{-valor} = 0.08183$, al realizar la comparación el valor p es mayor que alfa: $0.08183 > 0.05$, esto indica cumple con el supuesto de Normalidad.

3.3.3.3. Prueba de independencia

Para verificar la independencia del error aleatorio se realizó el gráfico de residuos y el orden de las observaciones.

Figura 13. Residuos contra orden del modelo 3



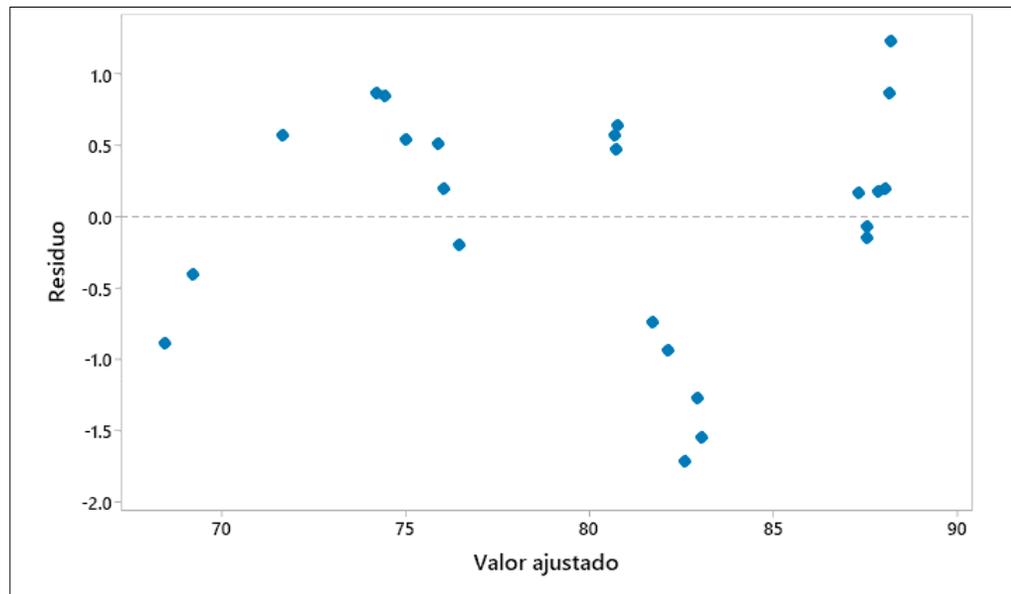
Fuente: elaboración propia.

Al igual que los modelo 1 y 2 no se puede concluir sobre el supuesto de independencia directamente con la gráfica. Se procede con la prueba de Durbin-Watson, como el valor de la muestra es el mismo que el modelo 1 al igual que los términos se compara el valor obtenido $D=1.54928$ de la tabla XV con un $D_L=1.18781$ y un $D_U=1.54639$, determinando que $D > D_U$. Los residuos son independientes, no existe correlación entre los residuos.

3.3.3.4. Prueba de homocedasticidad

Al realizó gráfico de Residuos vs. Ajuste y revisar la tendencia de los puntos se puede decir los errores tienen homocedasticidad o que las varianzas son iguales.

Figura 14. Residuos contra valor ajustado del modelo 3



Fuente: elaboración propia.

Se utilizó la prueba de Breush-Pagan con la hipótesis nula H_0 : los errores tienen varianza constante. El valor-p obtenido fue de 0.964 el cuál es mayor que el nivel de significación $\alpha=0.05$, no se puede rechazar la hipótesis nula lo que indica que los errores tienen varianza constante.

3.3.4. Resumen

En la siguiente tabla se resumen las pruebas aplicadas y si fueron superadas en cada modelo.

Tabla XVII. Resultados de las pruebas de bondad de ajuste y supuestos

Núm	Modelo	R ² ajustado en %	Normalidad	Independencia	Homocedasticidad	Mejor modelo
1	$B_x = -0.10306f - 0.2730A_t + 106.60$	97.26	no	si	si	no
2	$B_x = -0.10208f_c - 0.1954A_t + 101.56$	98.35	si	no	si	no
3	$B_x = 0.000178f_c^2 - 0.1558f_c + 93.837$	98.22	si	si	si	si

Fuente: elaboración propia.

De los resultados de la tabla se determina que el mejor modelo es el de orden dos al cumplir con todos los supuestos y tener un buen coeficiente de determinación.

El modelo estadístico que relaciona las variables que intervienen en el proceso de cristalización, utilizando los criterios del análisis de regresión y así mejorar la calidad en el tamaño y consistencia del grano de azúcar es:

$$B_x = 0.000178f_c^2 - 0.1558f_c + 93.837$$

4. DISCUSIÓN DE RESULTADOS

Authomathika (2021) indica en el manual del medidor automático μ -ICC 2.45 marca Promtec que cuentan con tres formas de mediciones las cuales son: fase, medición de nivel y atenuación, en los ingenios de azúcar la que mejor se adapta es por fase, así mismo establece que para la calibración, es necesario tomar muestras durante todas las etapas de la templa y evaluarlas en el laboratorio. Los datos que se deben observar en cada muestra son: fecha y hora, valor principal mostrado en ° brix, cambio de fase, temperatura, desplazamiento de fase compensado y atenuación.

La primera parte de la investigación se enfocó en identificar la relación de las variables que afectan en la cristalización de la azúcar determinada por la concentración de sacarosa medida en grados brix para cada etapa de las templeas de cocimiento con una muestra de 24 datos para cada variable regresora y la de respuesta. Todos estos datos cumplen con lo especificado en el manual del medidor automático.

Con esta información se realizó un gráfico de matriz como un análisis preliminar para la muestra de las variables regresoras fase, temperatura, atenuación, fase compensada y la variable de respuesta brix, con esta información se determina la relación de linealidad entre cada una de las regresoras con la de respuesta.

Con las variables definidas se aplicó un análisis de varianza con las cuatro variables regresoras, del cual se realiza una prueba de contrastes identificando cuales influyen de manera significativa en el modelo de regresión lineal múltiple,

identificando que las variables fase compensada es una de la que más aporta seguida de la atenuación. Sin embargo, fue necesario realizar un diagnóstico de multicolinealidad para verificar el factor de inflación de la varianza.

El coeficiente de determinación para este primer análisis indica que existe un buen ajuste del modelo, pero este no toma en cuenta la multicolinealidad de las variables, por lo que su resultado no es confiable. En la tabla V se muestran los valores de FIV para cada coeficiente de las variables determinando que la fase y fase compuesta inflan la varianza, por lo que se determinó quitar una de las dos del modelo, esto se confirma con lo expuesto por Montgomery *et al.* (2006). Al revisar la estructura de la variable fase compensada se encontró que está formada de la temperatura y la fase, por esto fue quitada del modelo.

Al realizar de nuevo el análisis de varianza sin la fase compensada, el modelo supera la prueba de no multicolinealidad, y la prueba de contraste indica que la temperatura no es significativa por lo que también se retira. El análisis de varianza para la fase y la atenuación muestra que los coeficientes son significativos el factor de inflación es adecuado y el coeficiente de determinación muy cercano a uno por lo que se determinó un primer modelo que cumple con la bondad de ajuste. Anderson *et al.* (2006) establece que estos coeficientes son los estimadores puntuales los que generan la ecuación del modelo de regresión.

Para lograr el segundo objetivo se estructuró el modelo con los coeficientes y el valor constante encontrados con el método de mínimos cuadrados, dado que se había eliminado la variable que corresponde a la fase compensada, para eliminar el problema de colinealidad se prueba otra combinación que incluye a la fase compuesta pero no a la variable fase.

Esta nueva combinación genera un nuevo modelo que incluye la fase compensada, la temperatura y la atenuación al realizar el análisis de multicolinealidad, prueba de ajuste y prueba de significancia de contrastes, dando como resultado que un modelo que incluye estas variables no cumple el criterio de multicolinealidad ni la significancia de contraste por lo que se procede a quitar la variable menos significativa que es la temperatura, la cual está relacionada con la fase compuesta.

El segundo modelo que incluye la fase compensada y la atenuación cumple con todos los criterios mostrados en el análisis de varianza de la tabla XII y se forma la ecuación de regresión lineal múltiple con los coeficientes de las variables y la constante encontradas con el método de mínimos cuadrados.

Se describe un tercer modelo en relación con información obtenida del fabricante del dispositivo de medición Promtec el cual indica que los parámetros de calibración del equipo pueden incluir una ecuación de segundo o tercer orden. La tabla XII muestra el porcentaje de contribución de cada variable regresora dentro del modelo el cual es de un 97.69 %.

Se inició el estudio del posible orden polinomial de la variable para ello se realiza un diagrama de dispersión Brix y fase compensada, este diagrama mostro que la relación entre las dos variables no es lineal, se define iniciar un modelo de orden dos para validar la significancia de contraste y bondad de ajuste con el coeficiente de variación.

En la tabla XIV se observan los resultados del análisis de varianza el cual cumple con la significancia de contrastes y el R cuadrado que muestra un excelente ajuste del 98.37 %.

Para el tercer objetivo sobre la validación de los modelos y establecer si representa adecuadamente el proceso de cristalización medido desde los grados Brix de la concentración de sacarosa, se realizaron las pruebas de hipótesis y comprobar los supuestos de normalidad, varianzas iguales o de homocedasticidad y de independencia. En esta parte coincide con lo expuesto por Navidi (2006) al indicar que se puede hacer un diagnóstico gráfico para validar los supuestos, pero se difiere ya que las gráficas de los residuos no fueron concluyentes y por esos se realizaron las pruebas de hipótesis.

En los tres modelos se realizó un análisis gráfico de normalidad con el QQ plot, el diagrama de orden contra residuos y el de valor ajustado vs residuos, en cada gráfico se interpreta el resultado en relación con el error del modelo y se afirma este resultado con la prueba de hipótesis en la prueba de Kolmogorov-Smirnov corregida por Lilliefors para datos menores de 50 y probar que el error tiene distribución normal. La prueba de Durbin-Watson para verificar la independencia de los residuos y la de Breush-Pagan para probar la homocedasticidad o las varianzas iguales en el error del modelo.

Para el modelo que incluye las variables regresoras fase y atenuación se observó que en el gráfico de normalidad algunos puntos no siguieron la recta, posiblemente por datos atípicos y la no normalidad de los residuos. Al aplicar la prueba corregida para muestras menores de 50 se rechaza la hipótesis nula y se establece que el modelo no cumple con el primer supuesto, se realiza una inspección de los datos ingresados para ver la existencia de un atípico el cual no se detectó. El modelo supera las pruebas de independencia y de homocedasticidad y su bondad de ajuste R cuadrado es de 97.26 %.

El modelo con las variables regresoras fase compensada y atenuación presentó un R cuadrado del 98.35 %, y al realizar el mismo análisis que el primer

modelo se encuentra que los residuos o error tienen una distribución normal al no poder rechazar la hipótesis nula. El modelo tiene homocedasticidad al verificarlo con la prueba de Breush-Pagan.

No se pudo concluir sobre la correlación en los errores de observación ya que el estadístico de Durbin-Watson se encuentra entre los límites de correlación y no correlación, por lo que no se pudo establecer la independencia de los datos del error.

Se analizó el modelo cuadrático para validar los supuestos de error y se encontró por la prueba de hipótesis de normalidad que los datos del error provienen de una distribución normal, asimismo mostro que los residuos son independientes y que existe homocedasticidad en los residuos por lo que tienen varianzas iguales.

Una de las características que debe tener cualquier modelo de regresión lineal múltiple de primer orden o de segundo orden es que deben de cumplir todos los supuestos. De no cumplir alguno se debe realizar un análisis exhaustivo para verificar las posibles causas sin embargo al tener tres modelos que describen el comportamiento del proceso, se puede elegir uno de los tres.

El primer modelo de fase y atenuación no cumplió la prueba de normalidad y no se encontró problemas de atípico por digitación incorrecta, se opta a indicar que no se incluye dentro de los mejores por problemas de normalidad.

En el segundo modelo no fue posible con la prueba de independencia comprobar si la variable fase compensada y atenuación están correlacionadas por lo que al igual que el modelo anterior no es posible indicar que es el modelo de mejor ajuste.

El modelo de una sola variable, fase compensada, se considera como un modelo de regresión lineal múltiple especial, al incluir una sola variable pero que produce un término cuadrático adicional, el cual para encontrar sus parámetros se utiliza la misma metodología de mínimos cuadrados. Este modelo superó todas las pruebas aunado a la información del manual del dispositivo que la ecuación de calibración puede ser una lineal o polinomial, se consideró el tercer modelo como el de mejor ajuste.

CONCLUSIONES

1. Se identificó que la variable que más influye en la cristalización de azúcar medida con la concentración de sacarosa en grados brix fue la fase compensada, la fase y la atenuación. La variable temperatura no es significativa para el modelo.
2. Se construyeron dos modelos de regresión lineal múltiple de orden uno con las variables regresoras fase y atenuación, fase compensada y atenuación, y uno de orden dos con la variable fase compensada. Los tres modelos cumplen con los criterios de bondad de ajuste, con el criterio de linealidad, el coeficiente de determinación ajustado, las pruebas de significancia y multicolinealidad.
3. Se validaron de forma gráfica y analítica con pruebas de hipótesis los supuestos del error de cada ecuación, con las pruebas de normalidad de la distribución de los errores, la homocedasticidad, aleatoriedad e independencia determinando que solo un modelo cumple con todos los supuestos.
4. Se construyó un modelo estadístico que relaciona las variables que intervienen en el proceso de cristalización en los tachos de refinería de un ingenio azucarero que contempla únicamente la variable regresora fase compensada, siendo el modelo de orden dos el que mejor predice los grados brix de la concentración de sacarosa que define cada etapa de la templa de producción de granos de azúcar.

RECOMENDACIONES

1. Se debe continuar con la identificación de la relación de las variables con una muestra mínima de 33 datos para cada variable del proceso, con el fin de poder cumplir con el nivel de confianza y con esto mejorar la predicción del modelo al tener un margen de error menor de 2 grados brix.
2. Se hace necesario construir un modelo de regresión para los distintos tachos de cocimiento ya que no se puede generalizar el modelo encontrado por que los datos son específicos, únicos e inherentes del tacho muestreado. Esto conlleva a realizar todo el análisis de varianza y las pruebas de bondad de ajuste con todas las variables.
3. Se deben validar los supuestos de error antes de calibrar el medidor automático dado que los modelos creados pueden cumplir con la bondad de ajuste, pero no necesariamente con la normalidad de la distribución del error, la homocedasticidad y la independencia.
4. Se sugiere tomar muestras de una templa completa tanto del medidor como de laboratorio en el mismo instante para evitar incertezas en las comparaciones de las medidas.

REFERENCIAS

1. Anderson, D., Sweeney, D., Williams, T., Camm, J., y Cochran, J. (2016). *Estadística para Negocios y Economía*. Mexico: Cengage Learning.
2. Arriza, A. J., Fernández, F., López, M., Muñoz, M., Pérez, S., y Sánchez, A. (2008). *Estadística Básica con R y R-Commander*. Cádiz: Servicios de Publicaciones de la Universidad de Cádiz.
3. Authomathika. (2021). *Manual μ -ICC 2.45 AIO device*. Recuperado de <https://www.authomathika.com.br/admin/public/img/manual-icc-2.pdf>
4. Corea, J. (2008). Instructivo de Operación Tachos Elaboración Masa de Tercera. Escuintla: Ingenio Santa Ana.
5. De la Garza, J. (1995). *Análisis de la información mercadológica a través de la estadística multivariante*. México, D.F.: Alhambra Mexicana.
6. Domínguez, G. (2013). *Normalización del tamaño de cristal de azúcar en el proceso azucarero a través del desarrollo sustentable aplicado al balance energético (Tesis Postgrado)*. Altamira: Instituto Politécnico Nacional.
7. Golato, M. (09 de 2014). *CÁTEDRA SISTEMAS DE CONTROL*. Obtenido de Universidad Nacional de Tucumán: <https://catedras.facet.unt.edu.ar/sistemasdecontrol/wp->

content/uploads/sites/101/2015/12/AP11-Control-de-Cocimientos-de-az%C3%BAcar-Ing-Golato-Sep-2014.pdf

8. Gutiérrez, H., y De la Vara, R. (2008). *Análisis y diseño de experimentos*. México, D.F.: McGRAW-HILL.
9. Hernández, F., & Mazo, M. (30 de octubre de 2020). *Pruebas de Homocedasticidad*. Obtenido de Modelos de Regresión con R : https://fhernanb.github.io/libro_regresion/index.html
10. Lind, D., Marchal, W., y Wathen, S. (2012). *Estadística aplicada a los negocios y la economía*. Mexico, D.F.: McGRAW-HILL.
11. Lohr, S. (2000). *Muestreo Diseño y Análisis*. México: Thomson .
12. Minitab. (2020). *Multicolinealidad en Regresión* . Obtenido de Soporte de Minitab 19: https://support.minitab.com/es-mx/minitab/19/help-and-how-to/statistical-modeling/regression/supporting-topics/model-assumptions/multicollinearity-in-regression/#fnsr_1
13. Montanero, J. (2008). *Análisis Multivariante [versión PDF]*. Cáceres: Universidad de Extremadura. Obtenido de http://matematicas.unex.es/~jmf/Archivos/ANALISIS_MULTIVARIANTE.pdf
14. Montgomery, D., Peck, E., y Vining, G. (2006). *Introducción al Análisis de Regresión Lineal*. México: CECSA.

15. Mulet-Hing, M., y Fernández-Salazar, R. (2016). Automatización del tacho cristizador de azúcar crudo del central Julio Antonio Mella. *SciELO - Scientific Electronic Library Online*, 13. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2224-61852016000200008
16. Navidi, W. (2006). *Estadística para ingenieros y científicos*. México, D.F.: McGRAW-HILL/INTERAMERICANA.
17. Rodríguez, M., Machado, W., y Villamarin, A. (18 de enero de 2019). Muestreo para el control de calidad en el proceso de elaboración de envases metálicos para alimentos. *Ingeniería Investigación y Tecnología*, 20, 9. doi:<https://doi.org/10.22201/fi.25940732e.2019.20n2.017>
18. Scheaffer, R., Mendenhall, W., y Ott, L. (1987). *Elementos de muestreo*. México: Grupo Editorial Iberoamérica.
19. Siegel, S., y Castellan, J. (1995). *Estadística no paramétrica aplicada a las ciencias de la conducta*. México: Trillas.
20. Valle, D. E. (2018). *Modelo de regresión lineal multivariado aplicado a la relación entre variables químicas del clinker y variables físicas del cemento. (Tesis de Postgrado)*. Guatemala: Universidad de San Carlos de Guatemala. Obtenido de <http://www.repositorio.usac.edu.gt/9588/1/Diego%20Enrique%20Valle%20Garc%C3%ADa.pdf>

21. Walpole, R., Myers, R., Myers, S., y Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias*. México: PEARSON EDUCACIÓN.
22. Webster, A. (2001). *Estadística aplicada a los negocios y economía*. Bogotá: McGRAW-HILL.

APÉNDICES

Apéndice 1. Toma de muestra para análisis de laboratorio



Fuente: [Fotografía de Luis Alvarado]. (Ingenio Santa Ana, Escuintla. 2021). Colección particular. Guatemala.

Apéndice 2. Interfase hombre-maquina HMI del equipo de medición



Fuente: [Fotografía de Luis Alvarado]. (Ingenio Santa Ana, Escuintla. 2021). Colección particular. Guatemala.

Apéndice 3. Equipo de laboratorio para medir la concentración



Fuente: [Fotografía de Luis Alvarado]. (Ingenio Santa Ana, Escuintla. 2021). Colección particular. Guatemala.

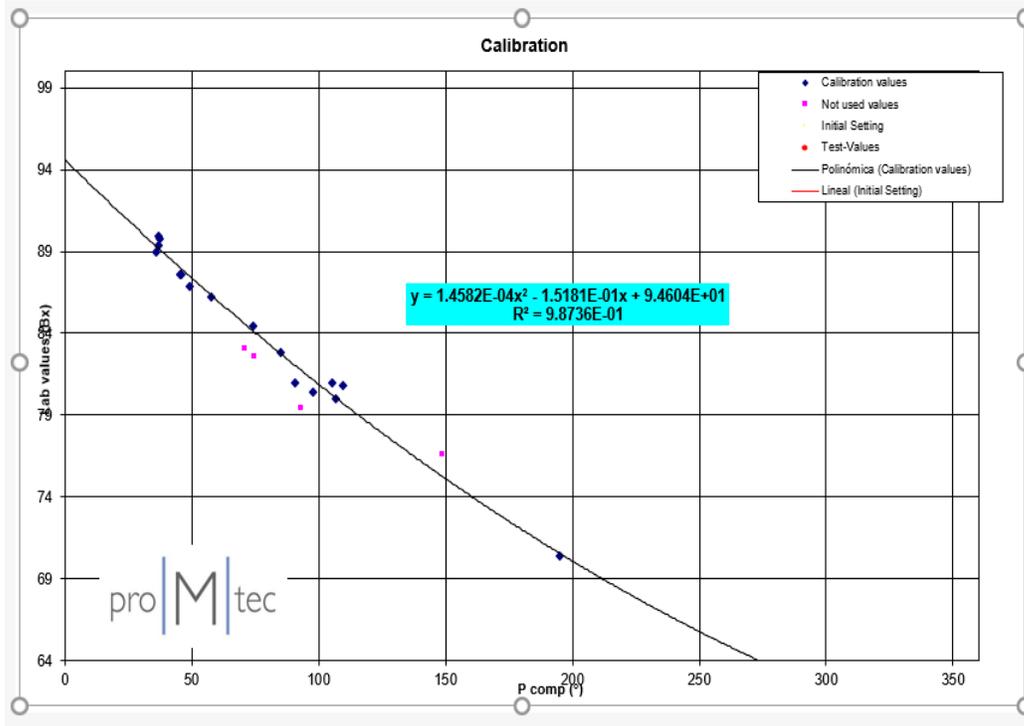
ANEXOS

Anexo 1. Datos de calibración tacho 3

FACTORY: INGENIO SANT ANA														proMtec					
PRODUCT: REFINERIA TACHO No.3																			
Date / Time		19/12/20 14:00																	
Initial setting		A0=	9.6239E+01	A2=	2.3442E-04	A4=	0.0000E+00	Tref=	65.0	Tk1=	2.0000E+00	Refcorr=	0.0						
Meas. Range		#VALOR!	9.6239E+01	A1=	-0.17770	A3=	0.0000E+00	A5=	0.0000E+00										
Date / Time																			
Actual setting		A0=	1.0000E+02	A2=	0.0000E+00	A4=	0.0000E+00	Tref=	65.0	Tk1=	2.0000E+00	Refcorr=	0.0						
Meas. Range		64.0	100.0	A1=	-1.0000E-01	A3=	0.0000E+00	A5=	0.0000E+00										
Found coeff														Tref=	65.0	Tk1=	2.0000E+00	Refcorr=	0.0
Meas. Range		49.6	93.4	A0=	93.4	A1=	-0.1216												
Correlation coe		R*	0.9827	DISPLAY				GRAPHIC VALUES				Actual setting		Found coefficients					
Sample	Date	Time	Brix	Phase(°)	P _{Comp} (%)	Temp(°C)	Atten(dB)	Brix Lab	P _{Comp} calc	Brix Lab	P _{Comp} calc	Stand. Dev=	29.88	Stand. Dev=	29.46				
												Calibr Instr	Lab-Calib Ins	Calibr Instr	Lab-Calib Ins				
1	semillamiento	08:55	80.11	120.12	109.43	70.30	59.62	80.80	109.52			89.05	-8.25	80.08	0.72				
2	Directa	10:05	88.36	55.64	45.56	70.00	49.92	87.60	45.64			95.44	-7.84	87.85	-0.25				
3	Concentracion	12:20	79.95	112.81	106.88	68.00	59.02	80.00	106.81			89.32	-9.32	80.41	-0.41				
4	semillamiento	12:40	79.85	116.54	105.09	70.70	58.47	81.00	105.14			89.49	-8.49	80.61	0.39				
5	Directa	13:05	84.31	81.35	74.59	68.40	54.52			82.60	74.55	92.55	-9.95	84.33	-1.73				
6	Directa	16:10	83.55	92.41	85.00	68.70	55.81	82.80	85.01			91.50	-8.70	83.06	-0.26				
7	Aprete Final	17:45	88.57	54.24	46.12	69.10	48.96	87.60	46.04			95.40	-7.80	87.80	-0.20				
8	Descarga	17:55	89.94	44.39	35.94	69.2	48.08	89	35.99			96.40	-7.40	89.02	-0.02				
9	Concentracion	18:35	75.13	150.92	148.69	66.1	60.14			76.6	148.72	85.13	-8.53	75.31	1.29				
10	semillamiento	18:50	82.58	102.03	90.77	70.60	57.99	81	90.83			90.92	-9.92	82.35	-1.35				
11	Directa	19:45	86.51	69.89	57.52	71.2	52.67	86.2	57.49			94.25	-8.05	86.41	-0.21				
12	Directa	22:40	87.64	55.24	52.2	68.1	49.76	86.9	49.04			95.10	-8.20	87.44	-0.54				
13	aprete previo	00:10	89.9	42.1	38.01	67.5	47.9	89.4	37.10			96.29	-6.89	88.89	0.51				
14	aprete fina	00:15	89.90	42.41	38.56	67.40	47.71	89.80	37.61			96.24	-6.44	88.83	0.97				
15	Descarga	00:30	90.08	41.34	37.88	67.3	47.15	89.9	36.74			96.33	-6.43	88.93	0.97				
16	Semillamiento	01:30	80.60	107.62	96.95	69.80	58.44	80.40	98.02			90.20	-9.80	81.48	-1.08				
17	Directa	02:05	84.99	85.50	66.22	72.20	51.81			83.10	71.10	92.89	-9.79	84.75	-1.65				
18	concentracion	04:05	70.45	196.28	195.04	65.60	57.59	70.40	195.08			80.49	-10.09	69.68	0.72				
19	semillamiento	04:25	80.77	101.62	93.97	69.20	59.09			79.40	93.22	90.68	-11.28	82.06	-2.66				
20	Directa	04:50	85.15	88.90	64.36	72.30	53.75	84.40	74.30			92.57	-8.17	84.36	0.04				
21									130.00			87.00	-87.00	77.59	-77.59				
22									130.00			87.00	-87.00	77.59	-77.59				
23									130.00			87.00	-87.00	77.59	-77.59				
24									130.00			87.00	-87.00	77.59	-77.59				
Test-Samples																			
1									130.00										
2									130.00										
3									130.00										

Fuente: Ingenio Santa Ana (2021). *Calibración de tachos de refinería*. Escuintla: Autor.

Anexo 2. **Modelo de regresión calculado por el software del fabricante**



Fuente: Ingenio Santa Ana (2021). *Calibración de tachos de refinería*. Escuintla: Autor.