



**UNIVERSIDAD POLITÉCNICA SALESIANA**  
**SEDE CUENCA**  
**CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA**  
**DE LOS RECURSOS NATURALES**

“ANÁLISIS DE LA TRANSFORMACIÓN DE DATOS ESPECTRALES OBTENIDOS  
POR ESPECTROSCOPÍA DE INFRARROJO POR TRANSFORMADA DE FOURIER  
PARA LA CUANTIFICACIÓN DE AZÚCARES”

Trabajo de titulación previo a la obtención del título  
de Ingeniero en Biotecnología de los Recursos Naturales

AUTORES: BRYAM PATRICIO QUIZHPE VÉLEZ

DAVID ALEJANDRO QUIZHPI UCHUARI

TUTOR: ING. PABLO WILSON ARÉVALO MOSCOSO, PhD.

Cuenca - Ecuador

2022

## **CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN**

Nosotros, Bryam Patricio Quizhpe Vélez con documento de identificación N° 0104824016 y David Alejandro Quizhpi Uchuari con documento de identificación N° 0107209777; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 19 de mayo del 2022

Atentamente,



---

Bryam Patricio Quizhpe Vélez

0104824016



---

David Alejandro Quizhpi Uchuari

0107209777

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE  
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Nosotros, Bryam Patricio Quizhpe Vélez con documento de identificación No. 0104824016 y David Alejandro Quizhpi Uchuari con documento de identificación No. 0107209777 expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Trabajo Experimental: “Análisis de la transformación de datos espectrales obtenidos por espectroscopía de infrarrojo por transformada de Fourier para la cuantificación de azúcares”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Biotecnología de los Recursos Naturales, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 19 de mayo del 2022

Atentamente,



---

Bryam Patricio Quizhpe Vélez

0104824016



---

David Alejandro Quizhpi Uchuari

0107209777

## CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Pablo Wilson Arévalo Moscoso con documento de identificación N° 0102156957, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: “ANÁLISIS DE LA TRANSFORMACIÓN DE DATOS ESPECTRALES OBTENIDOS POR ESPECTROSCOPIA DE INFRARROJO POR TRANSFORMADA DE FOURIER PARA LA CUANTIFICACIÓN DE AZÚCARES”, realizado por Bryam Patricio Quizhpe Vélez con documento de identificación N° 0104824016 y por David Alejandro Quizhpi Uchuari con documento de identificación N° 0107209777, obteniendo como resultado final el trabajo de titulación bajo la opción Trabajo Experimental que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 19 de mayo del 2022

Atentamente,



Ing. Pablo Wilson Arévalo Moscoso, PhD.

0102156957

## **DEDICATORIA**

La presente Tesis, va dedicada a mis padres, Carlos Quizhpi y Sandra Uchuari, por todo el apoyo y la confianza que han derramado en mi proceso universitario, de igual manera a mis hermanos y mi ángel Axel que me cuida desde el cielo y que me acompaño en todo mi proceso universitario.

A mi Dios padre, por llenarme de fe y motivación para seguir adelante y superar cada obstáculo.

## **DEDICATORIA**

La presente tesis va dedicada a nuestro Dios Padre y la Virgen Auxiliadora que me cuidan y cuidan cada una de los pasos que doy a diario, va dedicado a cada una de las personas que me creyeron en mi firmemente, entre ellos están mi familia Paulina, Patricio, Fanny, Viviana, Carmita, Pablo, Dani, Pablo que con su apoyo en todas las formas posibles me han ayudado a seguir siempre adelante y nunca desfallecer, que me ha acompañado desde el día uno hasta ahora, que me han dedicado tiempo y paciencia para poder culminar este proyecto de tesis, también va a dedicada a mis dos hijos que Dios padre me regalo Nicolás y Noelia que todo lo que hago es por y para ellos, son mi motor y mi ganas de superarme en las diferentes etapas de vida. Y a mi tía Magola que está en cielo cuidándome, apoyándome y sé que estará feliz, muy feliz viéndome culminar esta etapa.

## **AGRADECIMIENTOS**

Un agradecimiento muy especial a nuestros docentes, los cuales con su sabiduría nos han ayudado a formarnos académicamente y adquirir conocimientos para llegar al final de nuestras carreras, gracias a la Dra. Inés Malo y al Dr. Pablo Arévalo que nos han brindado tiempo y paciencia en todo el proceso de nuestro trabajo de investigación final.

Gracias a la Universidad Politécnica Salesiana, que ha sido nuestro lugar de formación académica y nos ha visto crecer como estudiantes y como personas, así mismo nos ha brindado las oportunidades para cumplir nuestros sueños.

## **AGRADECIMIENTO**

A mis padres, Carlos Quizhpi y Sandra Uchuari, por todos el apoyo y los esfuerzos realizados para conseguir mi formación académica. Gracias a ellos por los esfuerzos económicos que han hecho para que yo pueda llegar a donde estoy.

Agradezco a Dios, por guiar mi camino y proteger cada uno de mis pasos, me ha llenado de fe y salud.

A mis compañeros, que han sido de gran ayuda y apoyo para mi formación académica.

A mi compañero de vida Axel, gracias por la compañía durante toda mi vida universitaria, le agradezco hoy y siempre por los buenos momentos.

## **AGRADECIMIENTO**

Agradezco a Dios por darme la oportunidad y los medios para que todo esto se culmine, al esfuerzo de mi madre, de mi esposa que cada día pensaron en mí, para que todo esto se cumpla, sin desfallecer un solo segundo, por tenerme la fe, que todo esto se cumpla.

A mis hijos por darme esas fuerzas diarias, esas baterías interminables que se llenan cada día cumplir nuevas metas.

A mis compañeros de la universidad que cada día nos apoyábamos de alguna u otra manera para todos salir adelante.

## ÍNDICE DE CONTENIDOS

DEDICATORIA	v
AGRADECIMIENTOS	vii
ÍNDICE DE CONTENIDOS	x
ÍNDICE DE FIGURAS	xiv
ÍNDICE DE TABLAS	xv
RESUMEN	xvi
ABSTRACT	xvii
CAPÍTULO 1 INTRODUCCIÓN	1
1.1 Antecedentes	3
Antecedentes Interacionales	3
Antecedentes Nacionales	8
1.2 Problema de investigación	9
1.3 Formulación del problema o pregunta de investigación	10
1.4 Objetivos	10
<b>1.4.1. Objetivo general</b>	10
<b>1.4.2. Objetivos específicos</b>	10
1.5 Hipótesis	10
1.6 Justificación	11
1.7 Limitaciones	11

CAPÍTULO 2 MARCO REFERENCIAL	13
2.1. Estado del Arte	14
2.2 Bases teóricas	15
<b>2.2.1 Carbohidratos: Generalidades</b>	15
<b>2.2.1.1 Clasificación de los carbohidratos</b>	16
2.2.2 Métodos para la cuantificación de azúcares	18
<b>2.2.2.1 Métodos convencionales</b>	18
<b>2.2.2.2 Espectroscopía infrarroja</b>	19
<b>2.2.2.2.1 Interferómetro de Michelson</b>	20
<b>2.2.2.2.3 Espectroscopía Infrarroja por transformada de Fourier (FTIR)</b>	20
<b>2.2.2.2.3.1 Tratamientos matemáticos</b>	23
CAPÍTULO 3 MARCO REFERENCIAL	25
3.1. Nivel de investigación	26
3.2. Diseño de investigación	26
3.3. Variables	26
3.4. Población y muestra	27
3.5. Técnicas e instrumentos de recolección de datos	27
3.6. Técnicas de procesamiento de datos	27
3.7. Protocolo	27
<b>3.7.1. Análisis de Pureza</b>	27
<b>3.7.2. Niveles de Azúcares</b>	28

3.8. Procedimiento	29
3.8.1. <i>Colecta de Blanco</i>	29
3.8.2 <i>Colecta de la Muestra</i>	29
3.8.3. <i>Límites de Exposición</i>	30
3.8.4. <i>Normalización de la Escala</i>	30
3.8.5. <i>Automatic Baseline Correct (Corrección de la línea Base)</i>	30
3.8.6. <i>Automatic Smooth (Suavizacion del Espectro)</i>	30
3.8.7. <i>Primera Derivada</i>	30
3.8.8. <i>Segunda Derivada</i>	30
3.8.9. <i>Descarga de datos</i>	31
3.9.10. <i>Análisis Estadístico</i>	31
3.9.10.1. <b>Depuración de datos</b>	31
3.10. Análisis de Componentes Principales	31
3.11. Regresión Lineal	32
3.11.1. <i>Test de Error cuadrático medio (MSE) de (training).</i>	32
CAPÍTULO 4. RESULTADOS	33
4.1. Obtención de espectros de los azúcares	34
4.2. Análisis de Componentes Principales	36
4.3. Desviación Estándar de Cada Componente	37
4.4. Componentes de Importancia	38
4.4. Contribución a la Varianza	40

4.5. Modelo Matemático	41
<b><i>4.5.2. Ajustar el Modelo</i></b>	41
4.6. Resultados de la Fructosa	42
4.7. Resultados de la Sacarosa	43
4.8. Resultados de la glucosa	44
DISCUSIONES	46
CONCLUSIONES	48
RECOMENDACIONES	49
BIBLIOGRAFÍA	50
ANEXOS	53

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Interferómetro de Michelson.....	20
<b>Figura 2.</b> El espectro de FTIR del agua pura. ....	21
<b>Figura 3.</b> Espectros de FTIR, a. Sacarosa, b. Pectina y C. almidón de maíz .....	22
<b>Figura 4.</b> Espectros FTIR de estándares de 4 tipos distintos de azúcares .....	23
<b>Figura 5.</b> Niveles de azúcares .....	28
<b>Figura 7.</b> Espectros para la obtención de datos de segunda derivada.....	35
<b>Figura 8.</b> Reducción de datos significativos para la interpretación de datos.....	36
<b>Figura 9.</b> Ingreso de la base de datos al software Rstudio.....	36
<b>Figura 10.</b> Desviación estándar de cada componente .....	38
<b>Figura 11.</b> Barras de los seis componentes principales .....	39
<b>Figura 12.</b> Contribución a la varianza .....	40

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Propiedades de los carbohidratos .....	16
<b>Tabla 2.</b> Clasificación de monosacáridos y oligosacáridos más relevantes.....	16
<b>Tabla 3.</b> Regiones espectrales.....	19
<b>Tabla 4.</b> Tratamientos matemáticos.....	24
<b>Tabla 5.</b> Concentración de azúcares .....	29
<b>Tabla 6.</b> Componentes principales .....	39
<b>Tabla 7.</b> Componentes principales de fructosa .....	43
<b>Tabla 8.</b> Componentes de la sacarosa .....	44
<b>Tabla 9.</b> Componentes principales de la glucosa .....	45

## RESUMEN

La presente investigación abarca el establecimiento de un modelo estandarizado que permita realizar análisis de azúcares mediante la transformada de Fourier sintetizando las gráficas con regresión multivariante, de manera eficiente para el uso en los Laboratorios de Ciencias de la Vida de la Universidad Politécnica Salesiana, optimizando los espectros de absorción de azúcares obtenidos por espectroscopía infrarroja para suavizar el espectro, por transformada de Fourier, a través de modelos matemáticos. Con esto se tiene datos con mayor precisión, la formulación de un modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica y la validación de la ecuación obtenida de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares, se deduce que para la mezcla del nivel sacarosa 3.29, glucosa 1.19, y para fructosa es de 1 lo cual indica que luego de emplear un número alto de variables que han sido utilizadas como predictores, al ajustar un modelo de regresión lineal por mínimos cuadrados se produce un número reducido de componentes principales que se pueden explicar cómo la mayor parte de la varianza de los datos. Y en cuanto a la validación la ecuación obtenida de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares, lo que permite concluir que el coeficiente  $R^2$ , que en este caso es 1, indica que el modelo es capaz de predecir azúcares de las observaciones con las que se trabajó. El valor  $pvalue = 2.22e-16$ , es un indicativo de una posible redundancia entre los predictores.

**Palabras clave:** Transformación de datos espectrales, espectroscopía de infrarrojo, transformada de Fourier, cuantificación de azúcares.

## ABSTRACT

The present investigation covers the establishment of a standardized model that allows sugar analysis to be carried out using the Fourier transform, synthesizing the graphs with multivariate regression, efficiently for use in the life sciences laboratories of the Universidad Politécnica Salesiana, optimizing the absorption spectra of sugars obtained by Infrared spectroscopy to smooth the spectrum, by Fourier transform, using mathematical models. With this the data has greater precision, the formulation of a mathematical model through a multivariate regression that allows the quantification of sugars in the logarithmic expression and the validation of the equation obtained from comparisons of sugar samples, through the predictive quadratic correlation coefficient for use in subsequent analyzes of sugars, it follows that for the mixture of the sucrose level 3.29, glucose 1.19, and for fructose it is 1, which indicates that after using a high number of variables that have been used as predictors, when adjusting a linear regression model by least squares, produces a small number of principal components that can be explained as most of the variance in the data. And in terms of validation, the equation obtained from comparisons of sugar samples, through the predictive quadratic coincidence coefficient for use in subsequent dees analyses, which allows us to conclude that the R2 coefficient, which in this case is 1, indicates that the model is capable of predicting sugar from the observations with which it worked. The value  $pvalue=2.22e-16$ , is an indication of a possible redundancy between the predictors.

**Key words:** Spectral data transformation, infrared spectroscopy, Fourier transform, sugar quantification

# **CAPÍTULO 1**

## **INTRODUCCIÓN**

La cuantificación de los azúcares en los alimentos suele estar sujeta a una serie de limitaciones operativas y de costes. En la química tradicional, el método más conocido para la determinación de los azúcares reductores es el método Fehling, que se limita a la cuantificación de los azúcares reductores. Con el desarrollo de la cromatografía líquida de alto rendimiento (HPLC), la cuantificación de azúcares por este método es ahora una prueba estándar reconocida por la Asociación de Comunidades Analíticas (AOAC International). Sin embargo, una de las limitaciones de este método es el elevado coste derivado del coste del propio instrumento y de los costes adicionales de laboratorio (Pérez, 2017).

Por ello, los investigadores de todo el mundo están buscando otras formas de analizar los azúcares de forma rápida, barata y precisa. Una alternativa a este problema es la espectroscopía infrarroja, que puede medir la absorción causada por los movimientos vibratorios de ciertos grupos químicos en la región infrarroja del espectro (Pérez, 2017).

La región infrarroja del espectro se encuentra en el rango  $12500-100\text{ cm}^{-1}$  que se divide en tres regiones: NIR (infrarrojo cercano), MIR (infrarrojo medio) y FIR (infrarrojo lejano) (Dufour 2009). La región NIR entre  $12500$  y  $4000\text{ cm}^{-1}$  es la primera región donde se observan las bandas de absorción correspondientes a las vibraciones moleculares. La región MIR ( $4000$  y  $400\text{ cm}^{-1}$ ) es la región más importante del espectro vibracional de las moléculas, ya que aquí se puede obtener toda la información sobre las moléculas orgánicas. El rango FIR entre  $400$  y  $100\text{ cm}^{-1}$  proporciona información conformacional sobre las estructuras (Dufour 2009). Dado que el examen de los espectros infrarrojos no permitía inicialmente interpretaciones muy precisas, se desarrollaron espectrofotómetros infrarrojos basados en la interferometría de Michelson. Michelson descubrió que la aplicación de las transformadas de Fourier a los espectros infrarrojos obtenidos con un interferómetro permitía una mejor resolución y una interpretación más fácil de las señales.

Tras este acontecimiento en el desarrollo de la espectroscopía de infrarrojos, el sistema pasó a conocerse como espectrometría FTIR (infrarrojo por transformada de Fourier) (Subramanian y Rodríguez-Saona 2011). Sobre la base de este concepto, se definió la gama del espectro infrarrojo en la que se pueden medir los hidratos de carbono. Este rango se sitúa entre 1500 y 800  $\text{cm}^{-1}$ . Diferentes carbohidratos tienen bandas de absorción en este rango, lo que hace que la generalización de los modelos sea problemática cuando se estudian alimentos con combinaciones de carbohidratos. Como los espectrofotómetros registran las bandas de absorción cada  $2\text{cm}^{-1}$ , el número de puntos del espectro es grande. Como cada uno de los puntos espectrales representa una variable, el número de variables presentes es grande.

El desarrollo de modelos para el análisis de azúcar en muestras líquidas mediante espectroscopía FTIR está muy extendido, pero los modelos desarrollados no siempre cumplen los parámetros de calidad del modelo. Los métodos mencionados anteriormente implican el preprocesamiento de los datos, la elaboración del mejor modelo de regresión que se ajuste a los datos y su posterior validación. Por esta razón, los modelos publicados no son en algunos casos concluyentes y no son reproducibles en otras condiciones de funcionamiento.

## **1.1 Antecedentes**

### **Antecedentes Internacionales**

Herrera et al. (2018) en su investigación titulada “Aplicación de espectrometría infrarroja por transformada de Fourier (FTIR) en el laboratorio de la Universidad Cooperativa de Colombia sede Villavicencio. [Tesis de pregrado]”. El objetivo era encontrar la firma espectral de los hormigones 2500, 3000 y 3500 (psi), para lo cual se utilizó un espectrómetro de infrarrojos (SCiO). Este sensor permitió encontrar la firma espectral de cada uno de ellos, Para obtenerla, fue necesario recoger un gran número de muestras para determinar una curva estadística que permitiera caracterizar cada uno de estos hormigones y encontrar una

herramienta de gran precisión y fácil manejo que permitiera ensayar en el campo y en el laboratorio el tipo de hormigón utilizado, y así tener un mayor porcentaje de fiabilidad en el desarrollo de los trabajos que habitualmente se realiza en la vida profesional.

García et al. (2018) en su artículo titulado “Estudio estructural de la humina y sus interacciones con ácidos húmicos mediante espectroscopía de infrarrojo medio con transformada de Fourier” el objetivo fue realizar un análisis estructural del humus de tres suelos colombianos mediante espectroscopía infrarroja por transformada de Fourier (FTIR) con reflexión total atenuada (ATR). Las muestras de suelo fueron pretratadas, caracterizadas fisicoquímicamente y se extrajeron de ellas las sustancias solubles, los ácidos fúlvicos, los ácidos húmicos y los complejos de arcilla Hu. Posteriormente, cada una de estas fracciones se analizó mediante el ATP y el coeficiente de correlación de Pearson, que refleja la similitud de estructura y composición entre las sustancias húmicas (Hu) y los ácidos húmicos. Las interacciones intermoleculares entre Hu y las moléculas orgánicas se investigaron mediante ATP y espectroscopía de derivación funcionalmente mejorada (FEDS), que reveló un desplazamiento de la banda vibracional del hidroxilo a una banda de onda inferior, lo que indica la presencia de enlaces de hidrógeno en la formación de agregados supramoleculares.

Vera (2019) en su investigación “Evaluación de la adulteración de leche con lactosuero por la técnica analítica: espectroscopía del infrarrojo con transformada de Fourier en la región media (FT-MIRS) y cercana (FT-NIRS)” El objetivo de este trabajo es investigar mediante dos aplicaciones de la espectroscopía infrarroja (infrarrojo cercano - FT-NIR e infrarrojo medio - FT-MIR) la búsqueda de zonas del espectro que indiquen la presencia de adulteración en términos cualitativos y/o cuantitativos. Al final del experimento no fue posible construir una curva de calibración con los datos obtenidos para la leche no adulterada y la leche adulterada en diferentes grados, pero se observaron los cambios causados por los diferentes contenidos de

suero en las muestras, lo que permitió concluir que el método de espectroscopía infrarroja es capaz de detectar la adulteración cuando está debidamente calibrado y que puede utilizarse como método preliminar de análisis de muestras para evitar el uso de otros métodos más fiables y que requieren más tiempo en los casos en que no sean absolutamente necesarios.

Orts (2021) en su investigación “Análisis de la leche mediante la espectroscopía de infrarrojos para estimar la presencia de mastitis en ganado caprino” Tuvo como objetivo encontrar un método para detectar la presencia de mastitis subclínica en cabras utilizando la espectroscopía de infrarrojos por transformada de Fourier (FT-MIR) en muestras de controles de leche oficiales. Para ello, se recogieron 540 muestras de leche de 180 cabras murcianas granadinas (360 muestras de glándula mamaria y 180 de leche de ordeño) procedentes de 7 explotaciones de la Comunidad Valenciana. Los animales fueron seleccionados para garantizar una alta variabilidad en el RCS, el número y el mes de lactancia. Las calibraciones realizadas para estimar los niveles de estos tres marcadores inflamatorios en muestras de leche de control mediante FT-MIR mostraron una baja precisión (coeficiente de determinación de 0.42 a 0.78; varianza explicada de 0.31 a 0.65). Esto sugiere que actualmente no es posible identificar a los animales afectados por la mastitis subclínica mediante análisis FT-MIR en muestras de leche de control.

Juli (2020) en su investigación “Capacidad de proceso en análisis de viscosidad cinemática por espectrometría de emisión atómica y FT-IR de aceite lubricante del motor de un tractor D8T para mantenimiento preventivo” Tuvo como objetivo principal mejorar la capacidad de fabricación del proceso de análisis de la viscosidad cinemática por espectrometría de infrarrojos por transformada de Fourier y espectrometría de emisión atómica en el aceite lubricante del motor del tractor D8T para su mantenimiento preventivo. Para analizar la evolución general de las propiedades fisicoquímicas, de los metales de desgaste y de las impurezas en función del

tiempo, se aplicó el método de la tendencia longitudinal utilizando equipos de medición de la viscosidad, la espectrometría de emisión atómica y la espectrometría infrarroja por transformada de Fourier, seguido de un análisis estadístico con el software Minitab. Los datos sobre las propiedades fisicoquímicas, los metales de desgaste y las impurezas del aceite lubricante se utilizaron para obtener y estimar los límites de control y el análisis del rendimiento, la evaluación y la recopilación de datos fuera de los parámetros de los límites de control para crear un historial de fiabilidad para el mantenimiento preventivo, y los límites de condena se analizaron para aplicar el plan de mantenimiento preventivo óptimo para el motor. Se puede concluir que la vida del motor del tractor D8T se prolonga mediante el control de las propiedades del aceite lubricante y las capacidades del proceso, y que el rendimiento del aceite lubricante se optimiza para un control límite de 250 horas.

Téllez (2019) en su investigación “Aplicaciones de la espectroscopía infrarroja en el análisis de alimentos” cuyo objetivo era analizar la espectroscopía del infrarrojo cercano NIR y del MIR medio como metodologías prometedoras. Se trata de métodos rápidos y no destructivos que tienen la gran ventaja de integrarse en los métodos de la química verde. Con el desarrollo de nuevos métodos e instrumentos quimiométricos, su potencial es cada vez mayor. Además, accesorios como las células de transmisión, los cristales ATR (reflexión total atenuada) y la microscopía FTIR están ampliando sus aplicaciones. En la industria alimentaria, las aplicaciones de la espectroscopía NIR y FTIR pueden dividirse en diferentes áreas. En el ámbito de la detección de fraudes en la industria alimentaria, una de las aplicaciones más típicas es la de garantizar la autenticidad de los alimentos, por ejemplo, la detección de adulteraciones en el aceite de oliva y en las hierbas aromáticas y especias para uso alimentario. Por otro lado, estas técnicas analíticas también ayudan a garantizar la calidad y seguridad de los alimentos y a prevenir incidentes como la crisis de la encefalopatía espongiforme bovina (EEB). Estos

métodos pueden utilizarse para detectar ciertos contaminantes en los alimentos, como los altos niveles de melamina en la leche infantil o el MCPD en el aceite de palma. En los últimos años se ha prestado cada vez más atención a garantizar la calidad de los alimentos, tanto por parte de los productores como de los consumidores. La espectroscopía infrarroja se presenta como un método rápido para investigar los parámetros de calidad de los alimentos. Además, este método analítico, que no es específico y proporciona información global sobre el alimento investigado, se utiliza para verificar la trazabilidad de los alimentos. Se han desarrollado varios métodos para determinar el origen geográfico de diferentes ingredientes alimentarios.

Díaz (2018) en su investigación “Validez del método de espectroscopía de infrarrojo por transformada de Fourier (FTIR) en la determinación de la cantidad de proteína verdadera de la leche de vaca (Majes - Arequipa, 2018), cuyo objetivo era establecer la validez de la espectroscopía infrarroja por transformada de Fourier (FTIR) para determinar el contenido de proteínas de la leche de vaca (Majes - Arequipa, 2018). Se aplicaron métodos de observación y encuesta, utilizando instrumentos como cuestionarios para los agricultores seleccionados. Las unidades de estudio consistieron en 10 muestras de leche cruda para el procedimiento de calibración, 20 muestras para el procedimiento de validación y 20 agricultores entrevistados. Los resultados muestran que, la espectroscopía infrarroja por transformada de Fourier (FTIR) es adecuada para la determinación del contenido de proteínas en la leche de vaca porque es lineal y  $R_2 = 0.9988$  cuando se ajusta a una línea recta. Es exacto y preciso, ya que alcanzó un coeficiente de variación del 5.2% en la prueba de repetibilidad y del 5.9% en la de reproducibilidad, ambos por debajo del parámetro de medición (6%). Y se trata de un método fiable, ya que la variación entre las mediciones realizadas por distintos analistas no es significativamente diferente de 0.03 y 0.04, y mediante la prueba de Fisher se obtuvo  $F_{cal} =$

3.6> Ftab 5.3. Por último, el contenido en proteínas de casi todas las muestras analizadas es inferior a la media teórica característica de la leche de buena calidad.

Ramírez (2019) en su investigación “Caracterización de muestras de textil para su reciclaje con espectroscopía de infrarrojo medio y módulo ATR” cuyo objetivo era comprobar si la espectroscopía del infrarrojo medio es un buen método para clasificar individualmente las distintas fibras textiles y facilitar así su tratamiento posterior. El método validado es una combinación de FTIR-ATR (para registrar los espectros de las muestras) y el procesamiento matemático de la matriz de espectros mediante el análisis de componentes principales (PCA). En este estudio en concreto, gracias a la colaboración con Fitex, se recogieron catálogos reales de las fibras textiles más utilizadas en el sector: algodón, lino, lana, seda, poliéster, poliamida, nylon y viscosa. Estas fibras se clasifican en naturales (algodón, lino, lana, seda) y sintéticas (poliéster, poliamida, nylon, viscosa). Para probar el método de análisis, se realizaron tres estudios: el primero con algodón y poliéster para comprobar si el método podía distinguir perfectamente entre estas dos fibras con características tan diferentes; el segundo con fibras naturales y el tercero para comprobar si podía distinguir entre diferentes fibras sintéticas. Los resultados de este primer estudio son muy positivos, ya que han permitido clasificar correctamente las diferentes fibras. Estos resultados sentaron las bases de un segundo estudio más amplio para clasificar las mezclas con diferentes composiciones de fibras que pudieran cuantificarse.

### **Antecedentes Nacionales**

Pérez (2017) en su investigación “Tratamiento matemático de señales espectrofotométricas infrarrojas para la cuantificación de azúcares” cuyo objetivo fue estudiar diferentes métodos matemáticos que permitan cuantificar la cantidad de Glucosa, Fructosa, Sacarosa y Lactosa, presente en muestras acuosas, a partir de espectros infrarrojos (FTIR). De todos los modelos

calculados, la lactosa no presentó un modelo efectivo para predecir su concentración. En los otros casos el mejor pretratamiento utilizado, fue la primera derivada, sin importar el filtro que se utilice, la selección de variables óptimas fue el Algoritmo Genético utilizando el método Mínimos cuadrados parciales (PLS) y los mejores modelos de regresión fueron PLS para la fructosa y PCR para la Glucosa y Sacarosa.

## **1.2 Problema de investigación**

La cuantificación de azúcares en un alimento comprende muchas limitaciones, se presentan varios problemas como el costo y los métodos utilizados. En la química convencional, uno de los procedimientos más conocidos para determinar azúcares, es el método de Fehling, que se limita únicamente a la cuantificación de azúcares reductores. Las complicaciones han motivado a los investigadores de todo el mundo a estudiar nuevos métodos para el análisis de estos azúcares de forma rápida, reproducible y económica. Una alternativa que ha presentado buenos resultados es la Espectroscopía infrarroja, método que permite medir la absorción inducida por los movimientos vibracionales producidos por grupos químicos específicos, en la región del espectro infrarrojo, inicialmente los espectros infrarrojos, no permitían interpretaciones muy exactas, ahora se desarrollaron espectrofotómetros infrarrojos que utilizan el interferómetro de Michelson, otorgando resultados con mayor exactitud (Nieto, 2018).

Además, al aplicar las transformadas de Fourier al espectro infrarrojo obtenido utilizando un interferómetro, las señales tienen mejor resolución y la interpretación de las mismas es más sencilla. Desde este acontecimiento, en el desarrollo de la espectroscopía infrarroja, se conoce a este sistema como espectrometría FTIR (*Fourier Transform Infrared*) (Mondragón, 2017). Con este concepto se ha identificado, la región del espectro infrarrojo, en la cual permite medir los hidratos de carbono. Para alcanzar el objetivo que es determinar un modelo matemático que

permita cuantificar la presencia de azúcares a partir de espectros infrarrojos FTIR, se ha considerado conveniente utilizar tipos de tratamientos matemáticos.

### **1.3 Formulación del problema o pregunta de investigación**

¿Para qué se genera el modelo matemático al analizar la transformación de datos espectrales obtenidos por espectroscopía de infrarrojo por transformada de Fourier de azúcares?

### **1.4 Objetivos**

#### **1.4.1. Objetivo general**

- Establecer un modelo estandarizado que permita realizar análisis de azúcares mediante la transformada de Fourier sintetizando las gráficas con regresión multivariante, de manera eficiente para el uso en los Laboratorios de Ciencias de la Vida de la Universidad Politécnica Salesiana.

#### **1.4.2. Objetivos específicos**

- Optimizar los espectros de absorción de azúcares obtenidos por espectroscopía Infrarroja para suavizar el espectro, por transformada de Fourier, a través de modelos matemáticos, con esto se tiene datos con mayor precisión.
- Formular un modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica.
- Validar la ecuación obtenida de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares.

### **1.5 Hipótesis**

Si se establece un modelo matemático estandarizado, entonces se puede realizar un análisis más eficiente.

## **1.6 Justificación**

El espectro de infrarrojo por transformada de Fourier (FTIR) es una transformación matemática de señales cuyo dominio describe la frecuencia de dichas señales. A diferencia de otras transformaciones típicamente usadas en la transformación de datos, por ejemplo, las denominadas “wavelets”. En el procesamiento de señales digitales se emplean transformaciones matemáticas de alta complejidad para, extraer información de un conjunto de datos resultantes de respuesta, en función a los parámetros de interés (Díaz, 2018).

La técnica de transformada de Fourier es una técnica de procesamiento de señales digitales, en el caso de señales de origen espectral, la transformada de Fourier es el método más popular y generalizado. La mejor estrategia consiste de tres pasos: primero; aplicar la transformación de datos punto a punto, segundo; analizar el efecto de la transformación contra características de interés y, por último; racionalizar la transformación o construir transformaciones más complejas mediante la combinación de transformaciones simples. El objetivo de este estudio fue establecer un modelo estandarizado que permita realizar análisis de azúcares mediante la transformada de Fourier sintetizando las gráficas con regresión multivariante, de manera eficiente para el uso en los Laboratorios de Ciencias de la Vida de la Universidad Politécnica Salesiana.

## **1.7 Limitaciones**

La presente investigación se lleva a cabo en los Laboratorios de Ciencias de la Vida de la Universidad Politécnica Salesiana sede Cuenca, es por ello que al necesitar que sea presencial la ejecución de los objetivos, dentro de las limitaciones consideradas para el desarrollo de este estudio son la pandemia por Covid-19 y el ingreso limitado a Laboratorios de Ciencias de la Vida.

Asimismo, se encontró que existe información limitada del tema en Ecuador debido a su especificidad, por ello no se encontraron gran variedad de antecedentes nacionales, sin embargo, las bases teóricas se apoyan en estudios de nivel internacional.

# **CAPÍTULO 2**

## **MARCO REFERENCIAL**

## 2.1. Estado del Arte

Existen diferentes estudios que permitieron cuantificar cantidades de azúcares como la glucosa, sacarosa, lactosa y fructosa, se menciona una cuantificación realizada a partir de espectros infrarrojos (FIR), para cuantificar la fructosa se ocupó un pre tratamiento denominado método de mínimos cuadrados parciales (PSL), siendo eficaz desde la primera derivada, sin importar el filtro ocupado y un pre tratamiento por PCR (Regresión de componentes principales) siendo efectiva para la cuantificación de la glucosa y sacarosa (Daza y Parra, 2021).

Otro estudio realizado para la cuantificación de fructosa y sacarosa en pulpa y jugo de uvilla (*Physallis peruviana*), mediante el método de espectroscopía infrarroja por transformadas de Fourier, con zona de infrarroja media ( $700-4000\text{ cm}^{-1}$ ), partió de un desarrollo de curvas de calibración y un modelo lineal, en donde se realizó la cuantificación de los azúcares con registros de espectros infrarrojos en modo de absorbancia, mediante diluciones seriadas de los azúcares, esto permitió lograr resultados que luego fueron comparables y validados mediante refractometría (Daza y Parra, 2021).

El mismo método se ocupó para el análisis cuantitativo en diluciones acuosas de azúcares, este estudio exploró el uso de la espectroscopía del infrarrojo cercano para la cuantificación de sacarosa y glucosa, se identificaron correlaciones de las concentraciones en disolución con señales espectroscópicas obtenidas, lo mismo se realizó para espectros después de aplicar la transformada de Fourier mediante un programa y utilizando hojas de cálculo en Excel, en todos los casos se identificaron longitudes de onda para lograr curvas de calibración con respuesta lineal, obteniendo respuestas adecuadas para la cuantificación de los azúcares en jugos de fruta comerciales y miel (Juli, 2020). La técnica FTIR se considera una técnica rápida y sencilla para

evaluar la autenticidad y clasificar los azúcares, como zumos y néctares y evitar posibles adulteraciones de los mismos (Mondragón, 2017).

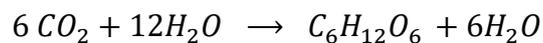
## 2.2 Bases teóricas

### 2.2.1 Carbohidratos: generalidades

Son biomoléculas compuestas por carbono, hidrógeno y oxígeno, también denominados glúcidos o glúcidos, su fórmula general es  $C_n(H_2O)_n$ , componentes esenciales de todos los organismos vivos, con el objetivo y función principal de proporcionar energía a través de la formación de glucosa (Téllez, 2019). Los carbohidratos son resultantes de la fotosíntesis, se forman gracias a la reacción del dióxido de carbono con agua, usando la clorofila como catalizador, la reacción se describe a continuación:



Son los compuestos principales que almacenan la energía del sol, la glucosa producida por las plantas gracias a la fotosíntesis, representa materia prima en gran cantidad para el desarrollo de varios alimentos, el dióxido de carbono se transforma en glucosa con el desprendimiento de oxígeno (Téllez, 2019). La reacción se describe a continuación:



Y mediante varias rutas bioquímicas, la glucosa da origen a diversas como; sacarosa y fructosa, también algunos polímeros como el almidón y la celulosa. La producción de energía que aportan los carbohidratos es de 4 kcal/g, en el cuerpo humano, esta energía se almacena en el hígado y en los músculos en forma de glucógeno. Por otro lado, los carbohidratos como; celulosa, agar y lignina no son digeribles por el cuerpo humano, pero son de gran importancia como parte de la fibra que se proporciona por los alimentos, además de ser parte de la pared

celular (Mondragón, 2017). En la tabla 1 a continuación, se describen las principales propiedades de los carbohidratos:

**Tabla 1.**

Propiedades de los carbohidratos

<b>Propiedad</b>	<b>Descripción</b>
<b>Solubilidad</b>	Interacción de la molécula de azúcar con el agua y otras moléculas, depende de los grupos hidroxilo.
<b>Cristalización</b>	La capacidad de formar cristales y de adaptar un estado amorfo de sobresaturación.
<b>Poder edulcorante</b>	Propiedad organoléptica.
<b>Poder reductor</b>	Poder de los átomos carbonílicos.
<b>Índice de refracción</b>	Energía luminosa que se transmite y refleja, propiedad que favorece determinar la concentración de azúcar.

*Nota. Tomado de Mondragón (2017)*

### 2.2.1.1 Clasificación de los carbohidratos

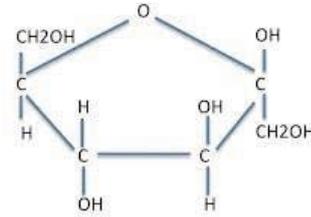
Se clasifican en tres grandes grupos, monosacáridos, que son los azúcares que no se hidrolizan, como ejemplo se puede mencionar a la glucosa, como segundo grupo de carbohidratos están los oligosacáridos, que son el producto de dos o más monosacáridos regularmente hasta 10, y por último los polisacáridos que están conformados por un gran número de monosacáridos, como el almidón (Juli, 2020). A continuación, en la tabla 2 se describe una breve clasificación de los monosacáridos:

**Tabla 2.**

*Clasificación de monosacáridos y oligosacáridos más relevantes*

<b>Monosacáridos</b>	<b>Descripción</b>	<b>Molécula</b>
Glucosa	Principal fuente de energía, se obtiene naturalmente en algunas frutas y verduras e industrialmente glucosa líquida y en polvo a partir de hidrólisis enzimática de almidón de cereales.	$  \begin{array}{c}  \text{O}=\text{C}-\text{H} \\    \\  \text{H}-\text{C}-\text{OH} \\    \\  \text{HO}-\text{C}-\text{H} \\    \\  \text{H}-\text{C}-\text{OH} \\    \\  \text{H}-\text{C}-\text{OH} \\    \\  \text{CH}_2\text{OH}  \end{array}  $

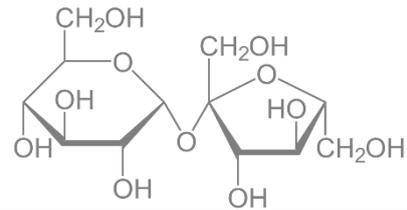
Fructosa De la familia de las cetosas, la fructosa es dos veces más dulce que la sacarosa, agente más utilizado en los productos alimenticios procesados.



**Oligosacárido** **Descripción**

**Molécula**

Sacarosa Azúcar no reductora, por carencia de grupo aldehído, integrada por una glucosa con un carbono aldehído unido al cetónico de la fructosa.




---

*Nota.* Tomado de Martín (2019).

El azúcar en las últimas décadas se ha vuelto un alimento fundamental en la dieta del ser humano, en la mayoría de países, podemos mencionar que la sacarosa o la habitual conocida azúcar de mesa, encabeza la lista de alimentos seguros emitido por la FDA (*Food and Drug Administration*), la elección de un tipo de azúcar para ser ocupado en la industria depende no únicamente del dulzor, también intervienen propiedades como el color, textura, porcentaje de humedad y además, el almacenaje y empaque que necesite el producto final (Martín, 2019).

Los mercados nacionales e internacionales buscan solucionar problemas relacionados al consumo excesivo del azúcar de caña, buscan una transformación de la industria convencional, aprovechando nuevas ideas y tecnologías que proporcionan las ciencias químicas y biotecnológicas, abriendo paso a nuevos desarrollos que permitan un uso total de la materia ocupada, en productos que generen sub productos y residuos que sean materia prima para nuevos productos o servicios, también llamado un desarrollo de ciclo cerrado (Mondragón, 2017).

## **2.2.2 Métodos para la cuantificación de azúcares**

### **2.2.2.1 Métodos convencionales**

La importancia de la cuantificación de azúcares radica en la seguridad de la composición dentro de un producto alimentario, incluyendo la trazabilidad desde su desarrollo, transporte y almacenamiento final. Para la determinación existen varios métodos, los tradicionales y no convencionales, un método muy ocupado es la refractometría, que consiste en una desviación de la trayectoria que se obtiene de un haz de radiación monocromática cuando atraviesa desde el vacío a otro material de densidad diferente (Téllez, 2019).

La refractometría presenta ventajas, es sencillo de desarrollar y además el procedimiento resulta económico, pero su rango de medición es corto y lectura es proporcionada por el lector lo que limita la exactitud, además, si las muestras no están distribuidas correctamente las lecturas analizadas no serán confiables.

Otro método que se describe es la polarimetría, tiene como fundamento principal un análisis químico, permite cuantificar el cambio que sufre un plano de luz polarizada al pasar por un medio transparente que está formado por sustancias ópticas activas (Ortiz, 2017), es un método de cuantificación simple y preciso, pero sus costos son elevados.

La espectrofotometría UV – Vis, es un método de alta sensibilidad y precisión para la cuantificación de azúcares, basa su fundamento en la absorción de la luz por la concentración de un compuesto, incide la luz mono cromática sobre un medio homogéneo, parte de la luz es absorbida y otra parte se transmite, la espectrofotometría es fácil de usar y necesita una mínima preparación de las muestras a analizar, el precio que requiere el equipo son altos por sus lámparas especiales (Téllez, 2019).

Existen métodos químicos para determinar azúcares, el más utilizado es el método Fehling, aunque se limita únicamente a la determinación de azúcares reductores, el método estándar más utilizado y reconocido hasta el momento por la asociación la de la comunidad analítica es la cromatografía de alta resolución (HPLC), pero los altos costos invertidos en este ensayo han llamado la atención de la comunidad para usar nuevas formas de cuantificar y determinar azúcares de manera exacta y económica, siendo una alternativa la espectroscopía infrarroja (Téllez, 2019).

### 2.2.2.2 Espectroscopía infrarroja

La base fundamental de la espectroscopía infrarroja es la medición de la absorción que es inducida por las vibraciones y movimientos producidos por los grupos químicos específicos en la región del espectro infrarrojo, el cual se encuentra entre  $12500 - 100 \text{ cm}^{-1}$ , la cual se divide en tres regiones, las cuales se describen en la tabla 3 a continuación:

**Tabla 3.**

#### *Regiones espectrales*

<b>Regiones</b>	<b>Descripción</b>	<b>Rangos</b>
NIR (Near – Infrared)	Se caracteriza por ser la primera región para observar bandas de absorción correspondientes a la vibración molecular	$12500 \text{ y } 4000 \text{ cm}^{-1}$
MIR (Mid – Infrared)	Es la región más importante del espectro vibracional de las moléculas, se puede obtener toda la información a moléculas orgánicas.	$4000 \text{ y } 400 \text{ cm}^{-1}$
FIT (Far – Infrared)	Esta región aporta con información sobre la conformación de las estructuras.	$400 \text{ y } 100 \text{ cm}^{-1}$

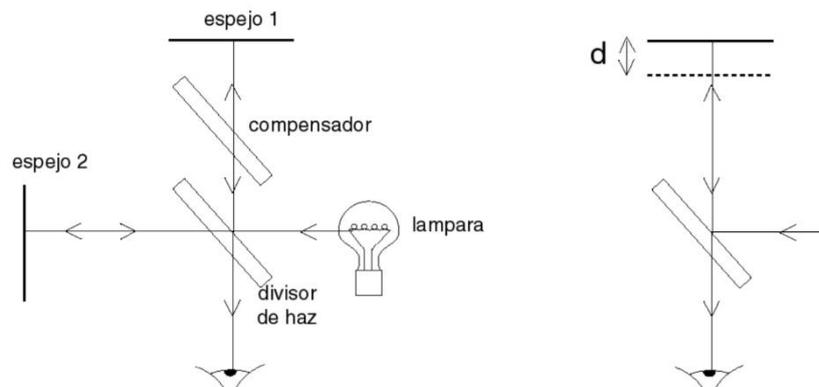
*Nota.* Los espectros infrarrojos no permiten interpretar resultados muy exactos, para esto se desarrollan, espectrofotómetros infrarrojos que usan el interferómetro de Michelson (Cruz, 2018).

### 2.2.2.2.1 Interferómetro de Michelson

El interferómetro es un instrumento que usa la luz para medir distancia, longitudes de onda e índices de refracción, Michelson describe el uso y la aplicación de métodos matemáticos, específicamente la transformada de Fourier al espectro obtenido, utilizando interferómetro las señales emitidas tienen una mejor resolución y la interpretación de los datos se vuelve más sencilla (Cruz, 2018), desde la incorporación de este modelo al método de espectroscopía infrarroja se conoce como espectroscopía FTIR (Song *et al.*, 2020). El interferómetro de Michelson se muestra en la figura 1, la luz se divide y los dos haces inciden en los espejos, la placa compensadora es necesaria para igualar los tamaños de los haces y puedan atravesar el ancho del vidrio, los haces serán recombinados.

**Figura 1.**

Interferómetro de Michelson



*Nota.* Elaboración propia.

### 2.2.2.3 Espectroscopía Infrarroja por transformada de Fourier (FTIR)

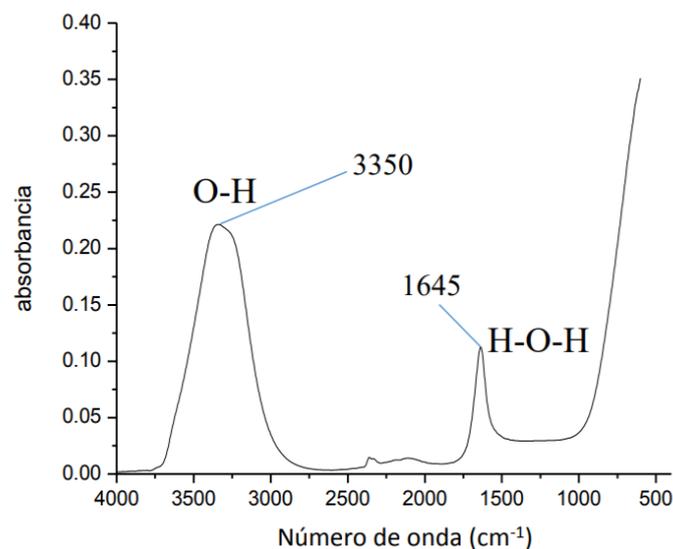
Existe un gran énfasis en el control de los ingredientes utilizados en empresas alimenticias, es de vital importancia conocer los diferentes ingredientes que intervienen en todo el proceso, sus propiedades químicas y hasta fisicoquímicas, ingredientes como los carbohidratos, lípidos,

proteínas y habitualmente el agua son controlados con intervalos que aseguren la calidad e inocuidad del producto que se ofrece, la modernidad en la que vivimos hace que los procesos de control sean mucho más fáciles, rápidos y confiables, la técnica FTIR permite un análisis completo a la hora de detectar componentes, cada uno con su respectivo espectro, visualizando los picos característicos de cada sustancia (Mondragón, 2017).

El espectro que más presente está dentro de las detecciones es el del agua, se obtienen de alimentos sólidos o semisólidos, tiene una tendencia marcada por los puentes de hidrógeno en su configuración, presenta dos picos notables uno en  $1645\text{ cm}^{-1}$  y otro alrededor de los  $3350\text{ cm}^{-1}$ , es un espectro muy sencillo, se puede visualizar en la figura 2, a continuación:

Figura 2.

El espectro de FTIR del agua pura. Nota: Es el espectro más conocido dentro de la espectroscopía infrarroja



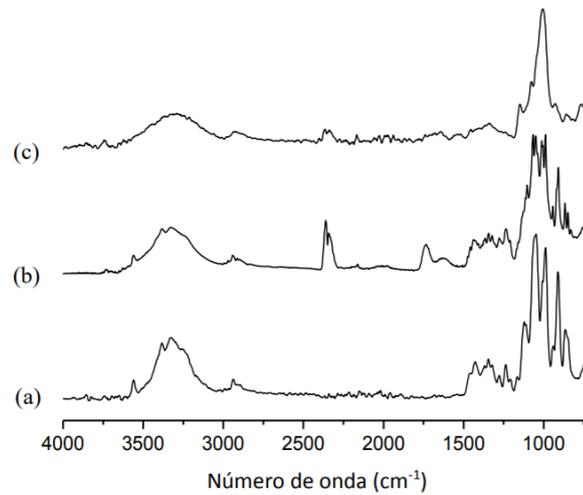
*Nota.* Tomado de Mondragón (2017)

El espectro de los carbohidratos por otro lado, presentan una característica particular, se visualiza un pico intenso, o dos en algunas ocasiones ubicados entre  $1200\text{ cm}^{-1}$  y  $900\text{ cm}^{-1}$ ,

dependiendo del carbohidrato el pico máximo se encuentra en una frecuencia específica (Téllez, 2019). A continuación, en la figura 3, se pueden observar los espectros FTIR de diferentes carbohidratos, con regiones marcadas entre  $1200$  y  $750\text{ cm}^{-1}$ .

**Figura 3.**

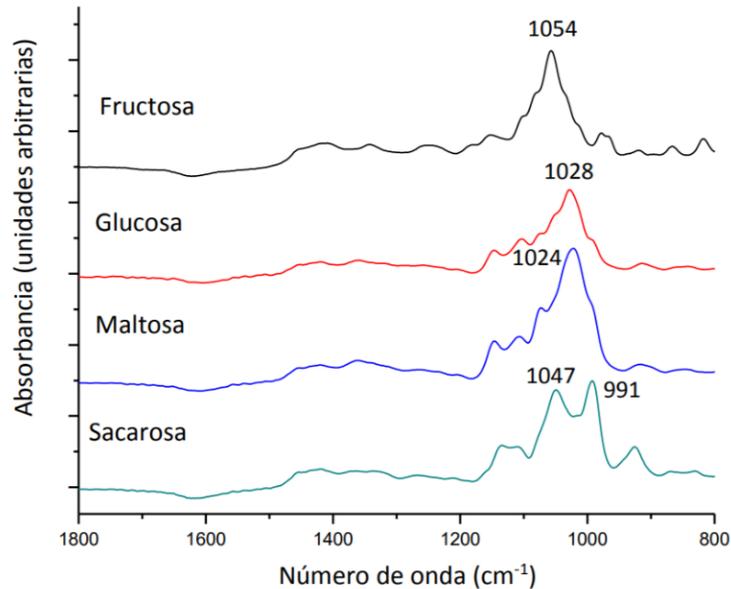
Espectros de FTIR, a. Sacarosa, b. Pectina y C. almidón de maíz



*Nota.* Tomado de Mondragón (2017)

**Figura 4.**

Espectros FTIR de estándares de 4 tipos distintos de azúcares, es posible diferenciar la posición de sus picos más intensos



*Nota.* Tomado de Mondragón (2017)

### **2.2.2.3.1 Tratamientos matemáticos**

Los tratamientos matemáticos son algoritmos y funciones que se usan para mejorar la interpretación de datos, los espectros obtenidos en FTIR son exclusivos para cada medición, es por esto que se realizan un tratamiento adicional antes de desarrollar el modelo matemático (Herrera et al., 2018).

Los tratamientos se deben realizar para poder interpretar los datos obtenidos de los espectros de una manera correcta, no está comprobado que un tratamiento sea mejor que otro, por lo que se debe realizar un estudio de los datos que se desean analizar y verificar cual sería el pre

tratamiento que mejores resultados obtendría, a continuación, se describen pre tratamientos matemáticos en la tabla 4:

**Tabla 4.**

*Tratamientos matemáticos*

<b>Tratamiento matemático</b>	<b>Descripción</b>
<b>Corrección de la línea base</b>	<p>Se realizan por lo general en software, consiste en analizar un ajuste cuadrático del espectro <math>Y(x)</math>, usando mínimos cuadrados.</p> $Y(x) \sim Y'(x) = ax^2 + bx + c$
<b>Suavización del espectro</b>	<p>Tiene su línea base en la eliminación de los datos de alta frecuencia del espectro, se utiliza el método de Savitzky – Golay, en donde se suavizan las señales, calculando una regresión polinomial, grado <math>n</math>, utilizando <math>n + 1</math> puntos en el espectro, con lo cual se tiene como resultados nuevos puntos.</p> $SF = \frac{K}{\frac{\%T * 2.302585}{100}}$ <p>En donde <math>K</math> es la frecuencia de la ecuación.</p>
<b>Normalización de la escala</b>	<p>En este tratamiento se relaciona la absorbancia a una longitud de onda determinada, los espectros van de un rango de 0 a 1, donde cero es el punto más bajo y uno el más alto.</p> $Normalizado = \frac{(Abs - AbsMin)}{(AbsMax - Abs Min)}$
<b>Derivadas</b>	<p>Transforman el espectro original, mostrando la información oculta en el espectro, la primera derivada obtiene un pico o valle, mientras que la segunda derivada, se observa invertida del espectro original, lo que ayuda a evidenciar los picos no definidos y con eso la información complementaria.</p>

*Nota.* Tomado de Díaz (2018).

## **CAPÍTULO 3**

# **MARCO REFERENCIAL**

### **3.1. Nivel de investigación**

La presente investigación será de tipo descriptiva dado que su base consta en el entendimiento de un problema a profundidad, para así poder describir los factores que intervienen en él. Según Hernández y Mendoza (2018), un estudio explicativo plantea entender la relación causal entre las variables de estudio, es decir, no solo se enfoca en describir un problema si no en entender las causas del mismo.

Además, será cuantitativo porque se recogen y analizan datos numéricos. Hernández y Mendoza (2018) indican que definen como aquel que utiliza la recusación y el análisis de datos para dar respuesta a las interrogantes del estudio y poner a prueba hipótesis determinadas y se basa en mediciones numéricas, cálculos y, a menudo, en el empleo de datos estadísticos para modelar de forma precisa el comportamiento de la población.

### **3.2. Diseño de investigación**

Como señalan Hernández y Mendoza (2018), los estudios en este contexto se denominan de tipo aplicado, es decir que se aplicarán los conocimientos aprendidos en la carrera para el logro de los propósitos planteados.

El diseño de la presente investigación será experimental, de acuerdo a Hernández y Mendoza (2018), está constituido por una metodología estadística que permite conocer y cuantificar las causas del efecto en el estudio que se realiza. Se caracteriza porque una o más variables relacionadas con el objetivo de la investigación son manipuladas de manera intencional y controlada con el fin de medir el efecto que tienen sobre otra variable de interés.

### **3.3. Variables**

- Variable Independiente: transformación de datos espectrales.
- Variable Dependiente: cuantificación de azúcares.

### **3.4. Población y muestra**

Conforme con Carrasco (2017) el conjunto de individuos de interés para el estudio, conforman la población y puede referirse a humanos, objetos o circunstancias que se deseen estudiar. Por tanto, la población estará conformada por una serie de muestras acuosas, con tres azúcares determinados (glucosa, sacarosa y fructosa).

Partiendo de esa premisa, la muestra se trata del subconjunto de la población, esto significa que es una representación de la población. Ahora bien, la muestra será de 10 muestras acuosas.

### **3.5. Técnicas e instrumentos de recolección de datos**

Por otro lado, como instrumento de recaudación de información, se utilizarán instrumentos electrónicos, los cuales según Hernández y Mendoza (2018) es cuando se mide la variable, a través del uso de un aparato que mide alguna cualidad como lo es la temperatura, medidas físicas, fuerza última de ruptura, entre otros; de la muestra de estudio. Por tanto, se utilizará un equipo espectrofotómetro infrarrojo THERMO Scientific.

### **3.6. Técnicas de procesamiento de datos**

Se obtendrán los datos del programa OMNIC, que es el controlador del equipo espectrofotómetro infrarrojo THERMO Scientific. Se recolectan los datos en una hoja de Excel y luego se analizarán con el fin de conocer cómo se comporta la cuantificación de los azúcares.

### **3.7. Protocolo**

#### **➤ Fase 1**

#### ***3.7.1. Análisis de Pureza***

Se procede a colocar en el espectrofotómetro muestras sólidas de fructosa, glucosa y sacarosa, con lo cual se comprueba que las muestras de azúcares a utilizar tengan la pureza adecuada.

La base de datos se obtendrá a partir de las diferentes muestras acuosas, con tres azúcares determinados, glucosa, sacarosa y fructosa; en el equipo espectrofotómetro infrarrojo THERMO Scientific, para lo cual se establecieron los siguientes aspectos para su formulación y elaboración.

### 3.7.2. Niveles de azúcares

Para la elaboración de las muestras se consideraron diferentes concentraciones a los mismos que se llaman niveles, para lo cual se han establecido cuatro, que se encuentran en la Figura 5, para lo cual se tomó en cuenta la disponibilidad de los mismos y la solubilidad de cada uno de los azúcares.

#### Figura 5.

Niveles de azúcares

<b>Nivel 1</b>	3.75 de glucosa	3.93 de fructosa	3.60 de sacarosa
<b>Nivel 2</b>	2.49 de glucosa	2.534 de fructosa	2.55 de sacarosa
<b>Nivel 3</b>	1.24 de glucosa	1.27 de fructosa	1.26 de sacarosa
<b>Nivel 4</b>	0.29 de glucosa	0.28 de fructosa	0.28 de sacarosa

*Nota.* Elaboración propia

Otro aspecto considerado para la formulación de las muestras es el azúcar presente en cada muestra, que va a depender del tipo de formulación, es decir, la concentración de la misma, la muestra 1, 2 y 3 tendrán la concentración de un solo azúcar que se llamará tipo pt (1), la muestra número 4, 5 y 6 es una combinación doble, la cual tendrá la presencia de dos concentraciones de azúcares con una relación de 0.5 entre las mismas, este tipo pt es el (2), en la muestra número 7 existe una combinación triple, en donde va a existir la concentración de tres azúcares con una fracción de 0.33 de cada una de ellas, el tipo pt va a ser llamado (3), en la muestra 8, 9 y 10 estarán presentes las tres concentraciones de azúcares, donde uno de ellos va a tener una presencia mayoritaria de azúcar en una fracción correspondiente a 0.625 y los otros dos una

fracción de 0.1875 cada uno, a este tipo pt es el (4), todo este procedimiento se hizo en cada uno de los niveles establecidos previamente en la figura 5, y se consideró dos réplicas de cada mezcla.

**Tabla 5.**

*Concentración de azúcares*

	<b>NIVEL 1</b>	<b>3.75 de glucosa</b>	<b>3.93 de fructosa</b>	<b>3.60 de sacarosa</b>
<b>Mezcla</b>	<b>tipo Pt</b>	<b>glucosa</b>	<b>Fructosa</b>	<b>sacarosa</b>
<b>1</b>	1	1	0	0
<b>2</b>	1	0	1	0
<b>3</b>	1	0	0	1
<b>4</b>	2	0.5	0.5	0
<b>5</b>	2	0.5	0	0.5
<b>6</b>	2	0	0.5	0.5
<b>7</b>	3	0.333	0.333	0.333
<b>8</b>	4	0.625	0.187	0.187
<b>9</b>	4	0.187	0.625	0.187
<b>10</b>	4	0.187	0.187	0.625

*Nota.* Elaboración propia

➤ **Fase 2**

**3.8. Procedimiento**

**3.8.1. Colecta de blanco**

Se colocó una muestra de una gota de agua Milli-Q® en el diamante, se presiona iniciar y el equipo realiza y grafica el espectro del blanco.

**3.8.2 Colecta de la muestra**

La siguiente actividad es colocar con un gotero la muestra del azúcar (glucosa, fructosa o sacarosa), se procede a correr la misma, y el equipo muestra el espectro junto con el blanco.

### ***3.8.3. Límites de exposición***

Seleccionando el espectro generado de la muestra analizada, a través de las opciones del programa (View), se colocan los límites a analizar del espectro, en este caso es de 1250 a 850.

### ***3.8.4. Normalización de la escala***

Luego de haber realizado los límites de exposición del espectro, se procede a normalizar la escala, a través de la opción (View) del programa OMNIC del equipo, que es de 0 a 1.

### ***3.8.5. Automatic Baseline Correct (Corrección de la línea Base)***

Luego de haber realizado la normalización de la escala del espectro se procesa a hacer la corrección de la línea base a través del programa OMNIC en la opción (Automatic Baseline Correct), esta entrega un nuevo espectro.

### ***3.8.6. Automatic Smooth (Suavización del Espectro)***

Luego de que se ha hecho el proceso de corrección de la línea base, se procede a realizar la suavización del espectro en el programa OMNIC a través de la opción (Automatic Smooth).

### ***3.8.7. Primera derivada***

Luego de la suavización del espectro se procede a realizar la primera derivada, a través de la opción (Analyze) del programa OMNIC, en donde se encuentra la opción de las derivadas, en la misma se selecciona (First) que permite obtener un espectro de la primera derivada.

### ***3.8.8. Segunda derivada***

Luego de haber obtenido la primera derivada se procede a realizar la obtención de la segunda derivada a través de la opción (Analyze), seleccionando la opción Derivative (Second) Segunda Derivada, el equipo da otro espectro junto con los demás ya obtenidos. Para visualizar todos los espectros, se utiliza la opción Stack spectra (apilar espectro).

### ***3.8.9. Descarga de datos***

Luego de haber realizado la segunda derivada se procede a extraer los datos de los espectros desde el mismo programa, a través de la alternativa (File, Download), los datos se descargan en formato Csv.text.

Este proceso se realiza con todas las 80 muestras que se generaron para levantar la información necesaria para proceder a realizar un análisis estadístico, a través del software RSTUDIO.

### ***3.9.10. Análisis estadístico***

#### **3.9.10.1. Depuración de datos**

Con la base obtenida de los espectros se procedió a determinar el rango de cada una de las longitudes de onda obtenidas ubicadas en las columnas de la hoja de la base de datos, aquellos rangos que no son representativos se procedió a no considerarlos, para lo cual se utilizó el programa Excel.

#### **3.9.10.2. División de datos para generación de modelo**

Luego de haber depurado la base de datos, se procedió a dividir la misma en tres bases de datos que corresponden a cada uno de los azúcares, lo cual se realizó en el programa EXCEL, obteniendo una base de datos en la que su primera columna correspondería a la concentración del azúcar y el resto de columnas representan la absorbancia en las diferentes longitudes de onda.

### **3.10. Análisis de componentes principales**

Para realizar el análisis de componentes principales, a partir de la base de datos se procedió a generar dos grupos, un grupo que contiene el 80% de los datos de la base que se lo llama entrenamiento y que ayuda a determinar los componentes principales, y ajustar el modelo, el

segundo grupo que corresponde al 20% (test) de la base de datos que se lo llama test que va a evaluar la capacidad predictiva del modelo generado.

A través de la obtención de los resultados del análisis se identificará cuáles son los componentes que se debería trabajar, complementariamente se recurre al *cross validation* para identificar el número óptimo de los componentes a utilizar.

### **3.11. Regresión Lineal**

#### ***3.11.1. Test de Error cuadrático medio (MSE) de (training).***

Luego de obtener el número correcto de componentes principales a utilizar para la predicción modelo, se determina el error cuadrático medio, todo esto con la base de datos entrenamiento. Complementariamente se va a utilizar la opción del programa conocido como la función *plsr()* del paquete *pls* para ajustar modelos por Partial Least Squares e incluye la posibilidad de recurrir a *cross validation* para identificar el número óptimo de componentes con el que se minimiza el MSE.

Este proceso se realizó para cada una de la base de datos de los azúcares, es decir glucosa, fructosa y sacarosa.

Se procede a evaluar la capacidad predictiva con el conjunto de datos llamado test y se determina el MSE, si este valor es igual o menor al obtenido en el modelo se puede validar el mismo.

# **CAPÍTULO 4.**

# **RESULTADOS**

#### **4.1. Obtención de espectros de los azúcares**

Previo a la obtención de espectros de las mezclas de los azúcares se hizo un análisis de pureza de las mismas, los resultados obtenidos establecen una pureza alrededor del 98%.

En la Figura 7, se puede observar un compilado de espectros que se generaron para obtener datos más precisos, de acuerdo a lo establecido en el protocolo; con color azul se puede observar el espectro del blanco que en nuestro caso fue agua Milli-Q®, el de color morado el espectro de la muestra de azúcar (glucosa). De acuerdo al protocolo se procede a la corrección de la línea base (color verde), para obtener un ajuste cuadrático del espectro, utilizando cuadrados lineales, encontrando la diferencia máxima entre todos los puntos del espectro lo que genera un subconjunto del espectro con la parte faltante de los picos principales.

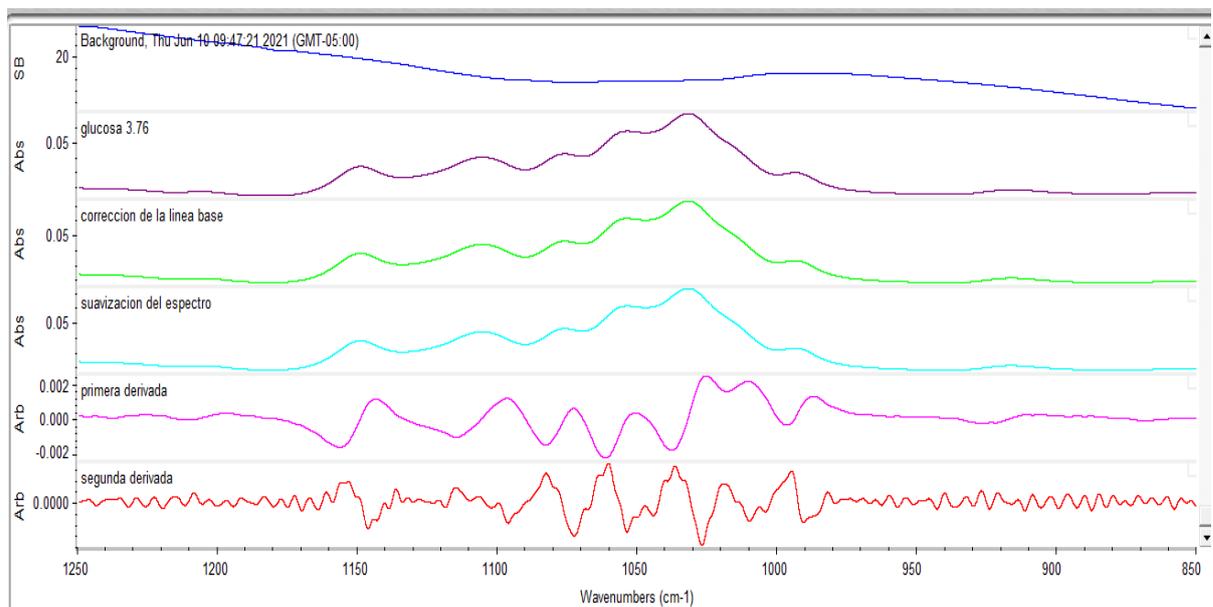
También se hizo la suavización del espectro que se visualiza en color celeste, que consiste en eliminar ruidos o datos de alta frecuencia del espectro que no son significativos o influyentes, a través de una regresión polinomial realizada por software OMNIC.

Estos espectros se expresan en una escala que viene por originalidad del programa, (eje y viene absorbancia entre valores mayores a cero), y para la obtención de la primera y segunda derivada, es necesario normalizar la escala, que consiste en variar la escala del eje y los mismos que van a estar entre valores de 0 a 1, luego de normalizado se procede a obtener la primera (color morado claro) y segunda derivada (color rojo) produciendo los espectros (color morado claro) la primera derivada al ser la pendiente de la curva se puede observar el espectro de la primera derivada que busca la pendiente de la curva, mientras que la segunda derivada visualizada en (color rojo), a través del programa OMNIC del equipo, este espectro de las derivadas permite identificar cierta información que no se puede observar en los espectros anteriores, evidenciando los picos que no están bien definidos, obteniendo información que

complementa los picos más complejos, siguiendo el mismo procedimiento para cada una de las muestras analizadas, por último se descargan los resultados de la segunda derivada para formar una base de datos.

### Figura 7

Espectros para la obtención de datos de segunda derivada.



Al obtener los datos de la segunda derivada de todas las muestras, se produce una base de datos conformada por 840 columnas que son las longitudes de onda con su respectiva absorbancia y 80 filas que corresponde a cada una de las muestras corridas.

Para poder hacer un análisis estadístico se hizo una depuración de datos basado en criterios establecidos en el protocolo, lo que produce una nueva base de datos con 80 filas y 265 columnas que representan las longitudes de onda, eliminando las que podrían alterar el resultado final, lo que se puede apreciar en la figura 8.

Figura 8.

Reducción de datos significativos para la interpretación de datos.

Nota. Elaboración propia.

## 4.2. Análisis de Componentes Principales

Figura 9.

Ingreso de la base de datos al software Rstudio.

	850.97199999999998	867.36469999999997	868.32899999999995	869.77539999999999	870.25750
G 1 (3,75) p	0.000000	0.000000	0.000000	0.000000	
G 1 (2,49) p	0.000000	0.000000	0.000000	0.000000	
G 1 (1,24)p	0.000000	0.000000	0.000000	0.000000	
G 1 (0,29)p	0.000000	0.000000	0.000000	0.000000	
G 2 (3,75)p	0.000000	0.000000	0.000000	0.000000	
G 2(2,49)p	0.000000	0.000000	0.000000	0.000000	
G 2 (1,24)p	0.000000	0.000000	0.000000	0.000000	
G 2(0,29)p	0.000000	0.000000	0.000000	0.000000	
G 1 (3,93)p	0.000000	0.000000	0.000000	0.000000	
G 1 (253)p	0.000000	0.000000	0.000000	0.000000	
G 1 (1,27)p	0.000000	0.000000	0.000000	0.000000	

Nota. Elaboración propia.

Se ingresaron los datos obtenidos de la segunda derivada de todas las muestras (base de datos 80 filas y 265 variables).

El análisis de componentes principales es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información, para este caso existe una muestra de 80 individuos, cada uno con 265 variables es decir, el espacio muestral tiene 265 dimensiones, el análisis de componente principal permitirá encontrar un número de factores adyacentes ( $F > \text{variables}$ ) que explican aproximadamente los mismos valores que las variables originales, el análisis de componentes principales no considera la variable de respuesta (columna 1), no se tiene en cuenta ya que el objetivo no es predecir la variable de respuesta (concentración de los azúcares), el método de componentes principales permite condensar la información procesada por múltiples variables en solo unos pocos componentes, ya que lo convierte en un método útil de aplicar previa utilización de otra técnica estadística como la regresión lineal, como es el caso de este trabajo.

Se procede a utilizar el software RSTUDIO para realizar el análisis de componentes principales para lo cual se utiliza el comando (prcomp), esta función permite ver las desviaciones típicas, la rotación, los centroides y las escalas, para este estudio se realiza una rotación con la función (head rotation)

#### **4.3. Desviación estándar de cada componente**

Como se puede observar en la figura 9, se indican los valores de la desviación estándar de los componentes, en la cual se puede apreciar que los componentes con valor mayor a 1, son desde el componente 1 al componente 27, tienen mayor desviación en relación a los demás, con esta tabla se tiene una idea clara tanto de la desviación estándar como de la varianza de los componentes.

**Figura 10.**

Desviación estándar de cada componente

```

> dim(respca$rotation)
[1] 264 80
> respca$sdev
[1] 7.891673e+00 6.155453e+00 4.864962e+00 4.442531e+00 3.834560e+00 3.202948e+00
[7] 2.952596e+00 2.847982e+00 2.734204e+00 2.599163e+00 2.451445e+00 2.389151e+00
[13] 2.326752e+00 2.211372e+00 2.189883e+00 2.074804e+00 2.025737e+00 1.875770e+00
[19] 1.744501e+00 1.655915e+00 1.614274e+00 1.359337e+00 1.287544e+00 1.138579e+00
[25] 1.110323e+00 1.075668e+00 1.031803e+00 9.759332e-01 9.506159e-01 9.322451e-01
[31] 8.790862e-01 8.465493e-01 8.209554e-01 7.843683e-01 7.823530e-01 7.078370e-01
[37] 6.928653e-01 6.434165e-01 6.193582e-01 5.699030e-01 4.569470e-01 4.233431e-01
[43] 4.213073e-01 3.087021e-01 2.873197e-01 1.929964e-01 1.147038e-01 1.371583e-07
[49] 7.979679e-08 6.150328e-08 3.508288e-08 3.504607e-15 2.207385e-15 1.910317e-15
[55] 1.477260e-15 1.345318e-15 1.334291e-15 1.053109e-15 9.345071e-16 7.682527e-16
[61] 7.011302e-16 5.705459e-16 5.705459e-16 5.705459e-16 5.705459e-16 5.705459e-16
[67] 5.705459e-16 5.705459e-16 5.705459e-16 5.705459e-16 5.705459e-16 5.705459e-16
[73] 5.705459e-16 5.705459e-16 5.705459e-16 5.705459e-16 5.343148e-16 3.483361e-16
[79] 3.378806e-16 2.761600e-16

```

*Nota.* Elaboración propia.

La misma indica la variabilidad de los componentes con respecto a la media, en donde se ve el peso de cada componente. El componente 1 tiene el 23% de proporción de varianza y el componente 2 tiene 14% de proporción de varianza, los dos primeros componentes indican el 37% del modelo.

#### 4.4. Componentes de Importancia

El Análisis de cada uno de los componentes, enseña la forma en que cada componente puede ayudar a interpretar la información que se recoge de cada componente, después de calcularlos, se procede a conocer la varianza explicada por cada uno de ellos, la proporción respecto al total y la proporción acumulada.

En el siguiente gráfico se puede apreciar los componentes principales con su desviación estándar, la proporción de varianza y la proporción acumulada de cada uno de los componentes, de la cual se considera para el estudio la proporción acumulada, en la cual se determina que

los seis primeros componentes explican el 63%, por lo tanto se puede decir que el resto de componentes van a influenciar en el 37%, es conveniente realizar el gráfico de barras en donde se observan los componentes principales a utilizar, el criterio aplicado para utilizar estos componentes es que, después del sexto componente principal la curva se aplana, alcanzando una explicación del 63%.

**Tabla 6.**

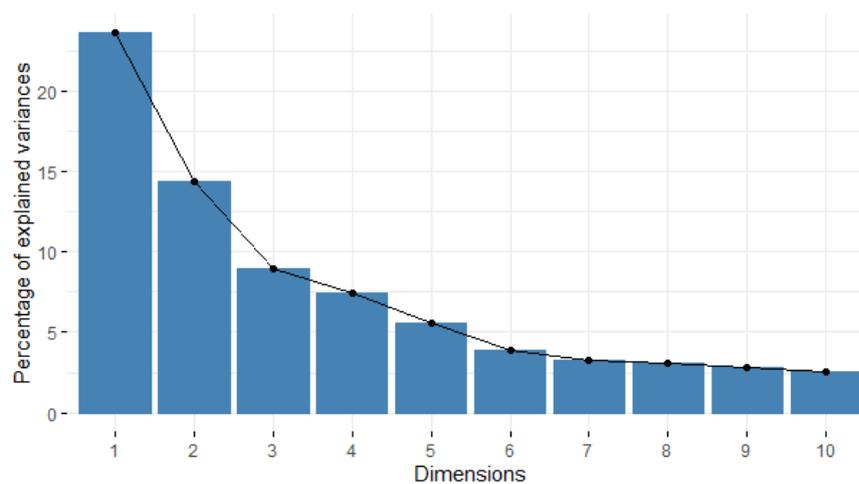
*Componentes principales*

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>
<b>Desviación estándar</b>	7.8917	6.1555	4.86496	4.44253	3.8346	3.20295
<b>Proporción de varianza</b>	0.2359	0.1435	0.08965	0.07476	0.0557	0.03886
<b>Proporción acumulada</b>	0.2359	0.3794	0.46908	0.54383	0.5995	0.63839

*Nota.* Elaboración propia.

**Figura 11.**

Barras de los seis componentes principales



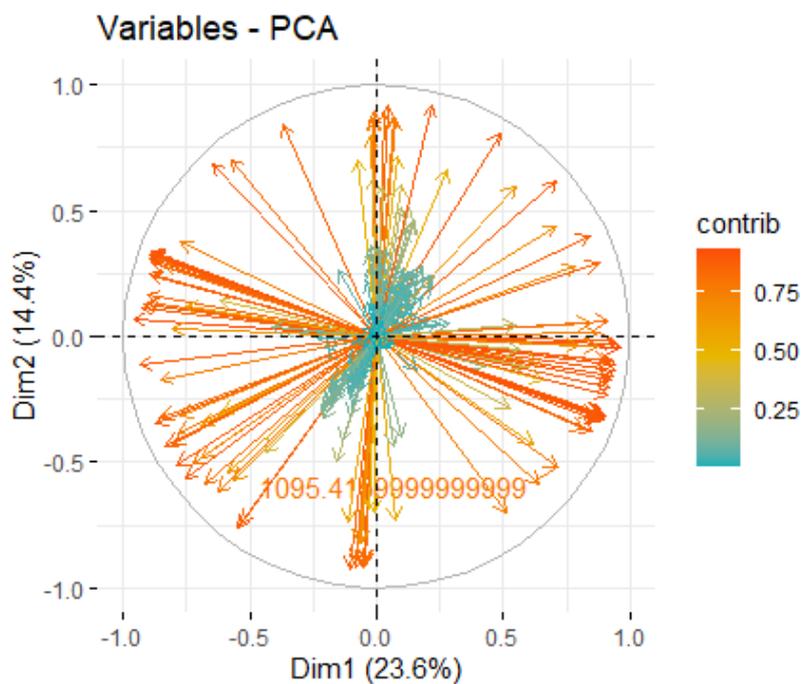
*Nota.* Elaboración propia.

#### 4.4. Contribución a la varianza

Se construye la gráfica (figura 12) de contribución a la varianza entre el componente 1 y 2 debido a que explica casi el 40% entre estos dos componentes, se aprecian las variables de color celeste, que son las que menos contribución tienen y las de color amarillo son las que más contribución tienen, en el mismo se puede observar representación de los individuos, donde se observa que los que tienen más contribución a la varianza se van alejando del centro, en razón de los componentes principales 1 y 2.

**Figura 12.**

Contribución a la varianza



*Nota.* Elaboración propia.

## **4.5. Modelo matemático**

### ***4.5.1. Regresión multivariante***

Una regresión con componentes principales se utiliza para realizar predicciones de una matriz (80 x 265) de variables dependientes normalmente distribuidas a través de la proyección sobre el subespacio óptimo generando una nueva matriz (80 x 16) de variables independientes, las mismas que serán utilizadas en un modelo de regresión lineal.

El método de regresión de componentes principales consiste en generar un modelo, de tal manera que con un número reducido de componentes se puede explicar la mayor parte de la varianza de los datos, en este estudio se dispone de un número elevado de variables, que en este caso es 265 que se pueden emplear como predictores, a pesar de ello un número alto de predictores no implica mucha relación, si estas variables están correlacionadas entre ellas la información es redundante, y además viola la condición de no colinealidad necesaria en la regresión por mínimos cuadrados, por lo tanto, se utilizó el análisis de componentes principales para eliminar información redundante empleando como predictores los componentes principales.

Para este trabajo, primero se procede a evaluar la capacidad de ajuste del modelo, para lo cual se dividen las observaciones disponibles en dos grupos, uno de entrenamiento para ajustar el modelo (80% de los datos) y uno de test (20% de los datos) para la validación del modelo. Los dos grupos de datos estaría conformada por 265 columnas, las mismas que corresponden a las longitudes de onda y una columna que es a la cual vamos a predecir.

### **4.5.2. Ajustar el modelo**

El grupo de datos para ajustar el modelo son ingresados, con el nombre entrenamiento, se realiza el cálculo de componentes principales, excluyendo la columna de la variable.

A través del comando `prcomp` con la base de datos se procede a hacer el cálculo de componentes principales, excluyendo la columna de la variable, en este caso (concentración de azúcares). Dando como resultado la tabla 7 de componentes principales.

Este mismo proceso se realizó en el apartado anterior, el estudio de la varianza explicada indica que el primer componente explica la mayor parte de la varianza acumulada (%23.59), decayendo drásticamente en los sucesivos componentes, por lo cual se consideran los seis primeros componentes para realizar la regresión lineal, entonces se aplican las funciones para realizar una regresión lineal en el programa RSTUDIO, para lo cual se utiliza el grupo de datos (entrenamiento) teniendo los siguientes resultados.

#### **4.6. Resultados de la fructosa**

Para la fructosa el valor de  $R_2$  es muy alto obteniendo 0.9948 lo que indica que el modelo es capaz de predecir con gran exactitud, además se puede observar el valor de *p-value*:  $< 2.2e-16$  lo que indica que el modelo en conjunto es significativo además se obtuvo el valor del cuadrado medio es de 0.01020937, para validar el análisis de regresión múltiple se procedió a realizar la corrida con el conjunto de datos (test) obteniendo un valor del error de cuadrado medio de 0.9491, este determina que el modelo está prediciendo de manera adecuada con lo cual se está validando que el modelo se está cumpliendo de una forma adecuada. En la tabla 7 se pueden observar los componentes principales con los que se trabajó para la fructosa.

**Tabla7.***Componentes principales de fructosa*

<b>FRUCTOSA</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>
<b>Desviación estándar</b>	179.5354	134.4206	120.9230	65.3172	44.0906	41.2170
<b>Proporción de varianza</b>	0.3909	0.2191	0.1773	0.0517	0.0236	0.0206
<b>Proporción acumulada</b>	0.3909	0.6100	0.7874	0.8391	0.8627	0.8833

*Nota.* Elaboración propia.

**4.7. Resultados de la sacarosa**

Para la Sacarosa el valor de  $R_2$  es muy alto obteniendo 0.3728 lo que indica que el modelo es capaz de predecir con gran exactitud, además se puede observar el valor de *p-value*:  $< 5.62e-13$  lo que indica que el modelo en conjunto es significativo además se obtuvo el valor del cuadrado medio de 0.06080495, para validar el análisis de regresión múltiple se procedió a realizar la corrida con el conjunto de datos (test) obteniendo un valor del error de cuadrado medio de 0.9363, este determina que el modelo está prediciendo de manera adecuada con lo cual se está validando que el modelo se está cumpliendo de una forma adecuada. En la tabla 8 se pueden observar los componentes principales de la sacarosa.

**Tabla 8***Componentes de la sacarosa*

<b>SACAROSA</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>
<b>Desviación estándar</b>	179.533	134.420	120.925	65.302	44.091	41.217
<b>Proporción de varianza</b>	0.3909	0.2191	0.1774	0.052	0.02358	0.0206
<b>Proporción acumulada</b>	0.3909	0.61	0.7874	0.83914	0.86272	0.8833

*Nota.* Elaboración propia.**4.8. Resultados de la glucosa**

Para la glucosa el valor de  $R_2$  es muy alto obteniendo 0.8602 lo que indica que el modelo es capaz de predecir con gran exactitud, además se puede observar el valor de *p-value*:  $< 2.161e-05$  lo que indica que el modelo en conjunto es significativo además se obtuvo el valor del cuadrado medio de 0.2700604, para validar el análisis de regresión múltiple se procedió a realizar la corrida con el conjunto de datos (test) obteniendo un valor del error de cuadrado medio de 0.9426, este determina que el modelo está prediciendo de manera adecuada con lo cual se está validando que el modelo se está cumpliendo de una forma adecuada. En la tabla 9 se pueden observar los componentes principales trabajados para la glucosa.

**Tabla 9***Componentes principales de la glucosa*

<b>GLUCOSA</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>
<b>Desviación estándar</b>	179.533	134.421	120.925	65.317	44.091	41.217
<b>Proporción de varianza</b>	0.3909	0.2191	0.1774	0.052	0.02358	0.0206
<b>Proporción acumulada</b>	0.3909	0.61	0.7874	0.83914	0.86272	0.8833

*Nota.* Elaboración propia.

## DISCUSIONES

La optimización de los espectros de absorción de azúcares obtenidos por espectroscopía Infrarroja por transformada de Fourier, a través de modelos matemáticos, en donde se evaluaron los rangos encargados de medir la intensidad de radiaciones absorbidas mediante sustancias en un rango entre 800 y 1250 en la glucosa, sacarosa y fructosa.

Mientras que Orts (2021) hace referencia a la espectroscopía de infrarrojos por transformada de Fourier (FT-MIR) en muestras de controles de leche oficiales. Para ello, se recogieron 540 muestras de leche de 180 cabras murcianas granadinas procedentes de 7 explotaciones de la Comunidad Valenciana. Las calibraciones realizadas para estimar los niveles de estos tres marcadores inflamatorios (lactoferrina) en muestras de leche de control mediante FT-MIR mostraron una baja precisión (coeficiente de determinación de 0.42 a 0.78; varianza explicada de 0.31 a 0.65). Esto sugiere que actualmente no es posible identificar a los animales afectados por la mastitis subclínica mediante análisis FT-MIR en muestras de leche de control.

Así mismo, la formulación de un modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica, en donde se mide la glucosa 3.60; 2.49; 1.24 y 0.29, sacarosa 3.6; 2.55; 1.27 y 0.28; fructosa 3.93; 2.53; 1.27 y 0.27 y las mezcla en los niveles 1, 2, 3 y 4, con valores de coeficiente de determinación múltiple, con relación a las bondades del modelo matemático. Por otra parte, Díaz (2018) los resultados muestran que: la espectroscopía infrarroja por transformada de Fourier (FTIR) es adecuada para la determinación del contenido de proteínas en la leche de vaca porque es lineal y  $R_2 = 0.9988$  cuando se ajusta a una línea recta. Es exacto y preciso, ya que alcanzó un

coeficiente de variación del 5.2% en la prueba de repetibilidad y del 5.9% en la de reproducibilidad, ambos por debajo del parámetro de medición (6%).

Y con respecto a la validación de la ecuación obtenida de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares, con el coeficiente  $R^2$  tiene un nivel óptimo de predicción mientras que Pérez (2017) estudiar diferentes métodos matemáticos que permitan cuantificar la cantidad de glucosa, fructosa, sacarosa, presente en muestras acuosas, a partir de espectros infrarrojos (FTIR). De todos los modelos calculados, la lactosa no presentó un modelo efectivo para predecir su concentración. En los otros casos el mejor pretratamiento utilizado, fue la primera derivada, sin importar el filtro que se utilice, la selección de variables óptimas fue el Algoritmo Genético utilizando el método Mínimos cuadrados parciales (PLS) y los mejores modelos de regresión fueron PLS para la fructosa y PCR (Regresión de Componentes Principales) para la glucosa y sacarosa.

## CONCLUSIONES

Con la optimización de los espectros de absorción de azúcares obtenidos por espectroscopía Infrarroja por transformada de Fourier, a través de modelos matemáticos, se obtiene datos con mayor precisión.

Es frecuente disponer de un número alto de variables que son empleadas como predictores, de los cuales no todos implican información interna, lo que indica que algunos valores son redundantes.

El Análisis de Componentes Principales elimina información redundante, al emplear como predictores los componentes principales, se puede realizar un modelo de regresión.

En cuanto a la validación de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares, lo que permite concluir que el coeficiente  $R_2$  para cada uno de las azúcares, que en este caso es  $>1$  (glucosa 0.8602, fructosa 0.9948, sacarosa 0.3728), indica que el modelo es capaz de predecir azúcares de las observaciones con las que se trabajó. El valor *pvalue*= (Glucosa  $< 2.161e-05$ , Fructosa  $< 2.2e-16$ , Sacarosa  $< 5.62e-13$ ), es un indicativo de una posible redundancia entre los predictores.

En función de la formulación del modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica, se deduce que para las cada una de los azúcares se obtiene: sacarosa 0.9363, glucosa 0.9426, y para fructosa es de 0.9491 lo cual indica que luego de emplear un número alto de variables que han sido utilizadas como predictores, al ajustar un modelo de regresión lineal por mínimos cuadrados se produce un número reducido de componentes principales que se pueden explicar como la mayor parte

de la varianza de los datos, indica que el modelo es útil, al aplicar componentes principales para predecir azúcares.

## **RECOMENDACIONES**

Se recomienda ampliar la investigación para cuantificar los espectros de absorción de azúcares, tanto glucosa, sacarosa y fructosa y comparar con otros métodos de FTIR para aumentar la precisión de los datos obtenidos.

Se recomienda recolectar una base de datos más corta y contar con instrumentos eficaces para evitar datos atípicos y garantizar la eficacia del modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica.

Se recomienda aplicar métodos correlativos estadísticos para contemplar los datos suministrados, la influencia de las observaciones, y el desempeño bajo un modelo de simulación de valores de las muestras de azúcares.

## BIBLIOGRAFÍA

- Carrasco, D. (2017). *Metodología de la Investigación Científica*. Lima: Editorial San Marcos.
- Cruz, H. (2018). *Interferencia y difracción II: interferómetro de Michelson*. México: Instituto de Ciencias Nucleares. Obtenido de [http://paginaspersonales.unam.mx/app/webroot/files/5268/interferencia\\_y\\_difraccion\\_II.pdf](http://paginaspersonales.unam.mx/app/webroot/files/5268/interferencia_y_difraccion_II.pdf)
- Daza, S., & Parra, G. (2021). Espectroscopia de infrarrojo con transformada de Fourier (ft-ir) para análisis de muestras de harina de trigo, fécula de maíz y almidón de yuca. *Limentech*, 5-16. Obtenido de [https://revistas.unipamplona.edu.co/ojs\\_viceinves/index.php/ALIMEN/article/view/4477/2644](https://revistas.unipamplona.edu.co/ojs_viceinves/index.php/ALIMEN/article/view/4477/2644)
- Díaz, C. (2018). *Validez del método de espectroscopía de infrarrojo por transformada de Fourier (FTIR) en la determinación de la cantidad de proteína verdadera de la leche de vaca Majes - Arequipa, 2018. [Tesis de maestría]*. Universidad Nacional de San Agustín de Arequipa. Perú. Obtenido de <http://repositorio.unsa.edu.pe/handle/UNSA/9383>
- Dufour, Éric. (2009) Principles of Infrared Spectroscopy Sun. Da-Wen. Infrared Spectroscopy for Food
- García, A., Combatt, E., & Palencia, M. (2018). Estudio estructural de la humina y sus interacciones con ácidos húmicos mediante espectroscopía de infrarrojo medio con transformada de Fourier. *Journal of Science with Technological Applications*, 4, 28-39. Obtenido de [https://www.researchgate.net/profile/Angelica-Garcia-Quintero/publication/332704124\\_Structural\\_study\\_of\\_humin\\_and\\_its\\_interaction\\_with\\_humic\\_acids\\_by\\_Fourier-transform\\_mid-infrared\\_spectroscopy/links/5cc46811a6fdcc1d49b463bf/Structural-study-of-humin-and-](https://www.researchgate.net/profile/Angelica-Garcia-Quintero/publication/332704124_Structural_study_of_humin_and_its_interaction_with_humic_acids_by_Fourier-transform_mid-infrared_spectroscopy/links/5cc46811a6fdcc1d49b463bf/Structural-study-of-humin-and-)
- Hernández, R., & Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. México: McGrawHill.

- Herrera, E., Benito, F., & Mancera, N. (2018). *Aplicación de espectrometría infrarroja por transformada de Fourier (FTIR) en el laboratorio de la Universidad Cooperativa de Colombia sede Villavicencio*. [Tesis de pregrado]. Universidad Cooperativa de Colombia, Villavicencio, Colombia. Obtenido de [https://repository.ucc.edu.co/bitstream/20.500.12494/6526/1/2018\\_aplicacion\\_espectrometria\\_infrarroja...pdf](https://repository.ucc.edu.co/bitstream/20.500.12494/6526/1/2018_aplicacion_espectrometria_infrarroja...pdf)
- Juli, C. (2020). *Capacidad de proceso en análisis de viscosidad cinemática por espectrometría de emisión atómica y FT-IR de aceite lubricante del motor de un tractor D8T para mantenimiento preventivo*. [Tesis de pregrado]. Universidad Nacional del Altiplano, Puno, Perú. Obtenido de <http://tesis.unap.edu.pe/handle/UNAP/13825>
- Martín, A. (2019). *Desarrollo de nuevas metodologías avanzadas para el análisis de oligosacáridos bioactivos*. [Tesis de maestría]. Universidad Complutense de Madrid, España. Obtenido de <https://eprints.ucm.es/id/eprint/56822/>
- Mondragón, P. (2017). *Espectroscopía de infrarrojo para todos*. Ciatej. Obtenido de [https://ciatej.mx/files/divulgacion/divulgacion\\_5a43b7c09fdc1.pdf](https://ciatej.mx/files/divulgacion/divulgacion_5a43b7c09fdc1.pdf)
- Nieto, M. (2018). *Caracterización estructural de materiales cristalinos: análisis de imágenes con resolución atómica mediante la transformada de Fourier*. [Tesis de pregrado]. Universidad de Valladolid, España. Obtenido de <https://uvadoc.uva.es/handle/10324/32057>
- Ortiz, A. (2017). *Sustitución del clarificante octapol con celite en los análisis de polarimetría de muestras obtenidas en el proceso de extracción de sacarosa en caña de azúcar*. [Tesis de pregrado]. Universidad Técnica del Norte, Ecuador. Obtenido de <http://repositorio.utn.edu.ec/handle/123456789/5967>
- Orts, Y. (2021). *Análisis de la leche mediante la espectroscopía de infrarrojos para estimar la presencia de mamitis en ganado caprino*. [Tesis de pregrado]. Universitat Politècnica de València, España. Obtenido de <https://riunet.upv.es/bitstream/handle/10251/173472/Orts%20-%20Analisis%20de%20la%20leche%20mediante%20la%20espectroscopia%20de%20infrarrojos%20para%20estimar%20la%20presencia%20d...pdf?sequence=2&isAllowed=y>

- Pérez, B. (2017). *Tratamiento matemático de señales espectrofotométricas infrarrojas para la cuantificación de azúcares. [Tesis de maestría]*. Universidad del Azuay, Cuenca, Ecuador. Obtenido de <https://dspace.uazuay.edu.ec/bitstream/datos/6854/1/12825.pdf>
- Ramírez, E. (2019). *Caracterización de muestras de textil para su reciclaje con espectroscopía de infrarrojo medio y módulo ATR. [Tesis de grado]*. Escola Politècnica Superior, España. Obtenido de <https://repositori.udl.cat/handle/10459.1/66740>
- Scientific, T. (2018). *Academia de espectroscopía FTIR*. Obtenido de Academia de espectroscopía FTIR: <https://www.thermofisher.com/ec/en/home/industrial/spectroscopy-elemental-isotope-analysis/spectroscopy-elemental-isotope-analysis-learning-center/molecular-spectroscopy-information/ftir-information.html>
- Song, Y., Cong, Yuanhua, Wang, B., & Zhang, N. (2020). Applications of Fourier transform infrared spectroscopy to pharmaceutical preparations. *Expert Opin Drug Deliv*, 17(4), 551-571. Obtenido de <https://pubmed.ncbi.nlm.nih.gov/32116058/>
- Subramanian, Anand y Luis Rodriguez-Sanoa. *Fourier Transform Infrared (FTIR) Spectroscopy*. New York-
- Téllez, C. (2019). *Aplicaciones de la espectroscopía infrarroja en el análisis de alimentos. [Tesis de pregrado]*. Universidad de Sevilla, España. Obtenido de <https://idus.us.es/handle/11441/91690>
- Vera, R. (2019). *Evaluación de la adulteración de leche con lactosuero por la técnica analítica: espectroscopía del infrarrojo con transformada de fourier en la región media (FT-MIRS) y cercana (FT-NIRS). [Tesis de pregrado]*. Fundación Universidad de América, Colombia. Obtenido de <http://repository.uamerica.edu.co/handle/20.500.11839/7380>

## ANEXOS

### Anexo 1. Operacionalización de las variables

Variables	Tipo de variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Escala de medición
Trasformación de datos espectrales obtenidos por espectroscopía de infrarrojo por transformada de Fourier	Independiente	La base fundamental de la espectroscopía infrarroja es la medición de la absorción que es inducida por las vibraciones y movimientos producidos por los grupos químicos específicos en la región del espectro infrarrojo (Cruz, 2018)	Es el método que se utilizará para medir la cantidad de azúcares en las muestras	Espectroscopía infrarroja  Tratamientos matemáticos	Equipo espectrofotómetro infrarrojo THERMO Scientific  Algoritmos y funciones	Razón
Cuantificación de azúcares	Dependiente	La importancia de la cuantificación de azúcares radica en la seguridad de la composición dentro de un producto alimentario, incluyendo la trazabilidad desde su desarrollo, transporte y almacenamiento final (Téllez, 2019).	Es medir en una muestra la cantidad de azúcares lo cual puede ser utilizado en diferentes alimentos	Azúcares a medir	<ul style="list-style-type: none"> <li>● Glucosa</li> <li>● Fructosa</li> <li>● Sacarosa</li> </ul>	Razón

## Anexo 2. Matriz de consistencia

Problema	Objetivos	Variables	Metodología
<p><b>General</b></p> <p>¿Para qué se genera el modelo matemático al analizar la transformación de datos espectrales obtenidos por espectroscopía de infrarrojo por transformada de Fourier de azúcares?</p> <p><b>Específicos</b></p> <ul style="list-style-type: none"> <li>• ¿Como Optimizar los espectros de absorción de azúcares obtenidos por espectroscopía Infrarroja para suavizar el espectro, por transformada de Fourier, a través de modelos matemáticos, con esto se tiene datos con mayor precisión?</li> <li>• ¿Cómo formular un modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica?</li> <li>• ¿Como validar la ecuación obtenida de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares?</li> </ul>	<p><b>General</b></p> <p>Establecer un modelo estandarizado que permita realizar análisis de azúcares mediante la transformada de Fourier sintetizando las gráficas con regresión multivariante, de manera eficiente para el uso en los Laboratorios de Ciencias de la Vida de la Universidad Politécnica Salesiana.</p> <p><b>Específicos</b></p> <ul style="list-style-type: none"> <li>• Optimizar los espectros de absorción de azúcares obtenidos por espectroscopía Infrarroja para suavizar el espectro, por transformada de Fourier, a través de modelos matemáticos, con esto se tiene datos con mayor precisión.</li> <li>• Formular un modelo matemático a través de una regresión multivariante que permita la cuantificación de azúcares en la expresión logarítmica.</li> <li>• Validar la ecuación obtenida de comparaciones de muestras de azúcares, a través del coeficiente de correlación cuadrático predictivo para el uso en análisis posteriores de azúcares.</li> </ul>	<p><b>Independiente</b></p> <p>Transformación de datos espectrales obtenidos por espectroscopía de infrarrojo por transformada de Fourier</p> <p><b>Dependiente</b></p> <p>Cuantificación de azúcares</p>	<p><b>Tipo de investigación</b></p> <p>Cuantitativa</p> <p><b>Nivel de la investigación</b></p> <p>Explicativa</p> <p><b>Población</b></p> <p>muestras acuosas, con 3 azúcares determinados (glucosa, sacarosa y fructosa)</p> <p><b>Muestra</b></p> <p>10 muestras acuosas</p>

Anexo 3. Instrumento de recolección de datos

**Ficha de observación**

Mezcla	Tipo Pt	Glucosa	Fructosa	Sacarosa	Gráfico
1					
2					
3					
4					
...					
9					
10					

Anexo 4. Recolección de datos en laboratorio

Figura 13.

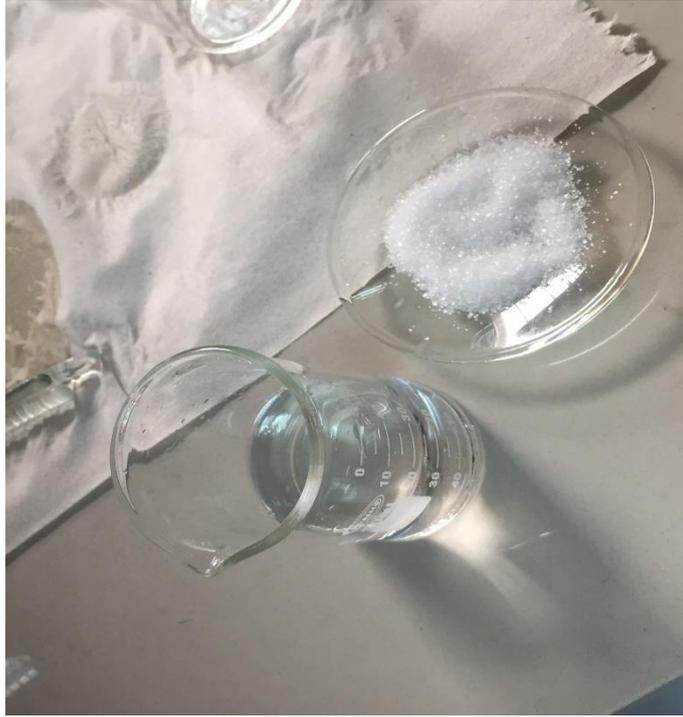
Glucosa, fructosa y sacarosa en cada uno de sus recipientes.



Nota. Elaboración propia

Figura 14.

Sacarosa siendo pesada en su estado reactivo.



Nota. Elaboración propia

Figura 15.

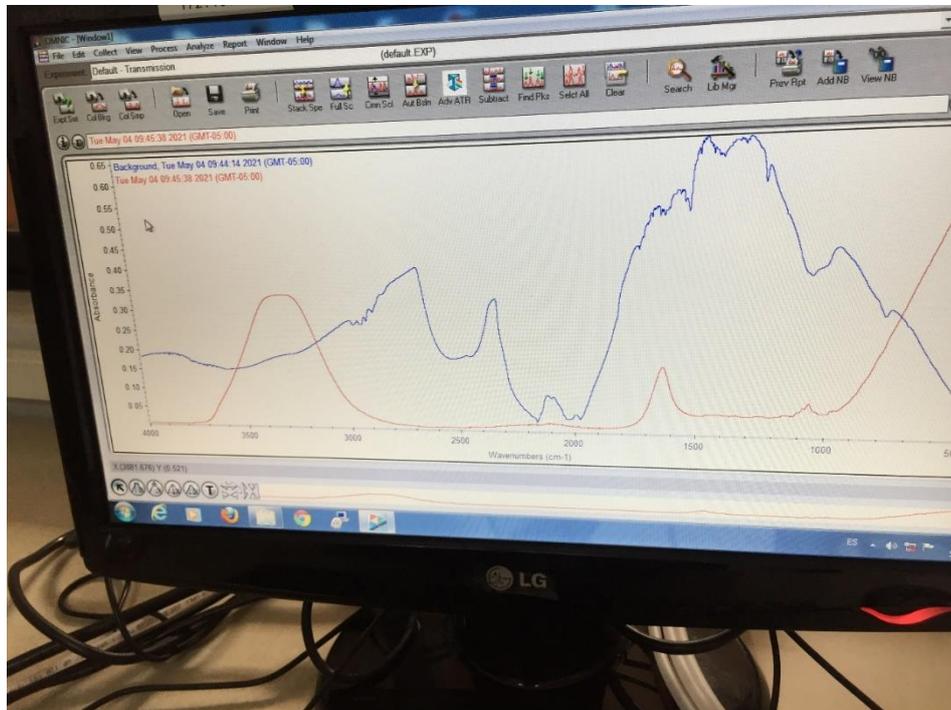
Glucosa 3.93 gramos en la balanza analítica.



Nota. Elaboración propia

Figura 16.

Visualización en el Software Omnic del espectro original de una muestra de azúcar junto con el blanco.



Nota. Elaboración propia