

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS  
DEPARTAMENTO DE POSTGRADO**

**PROYECTO DE TITULACIÓN**

**PREVIO A LA OBTENCIÓN DEL TÍTULO DE:**

**“MAGÍSTER EN ESTADÍSTICA APLICADA”**

**TEMA:**

**ESTUDIO MULTIVARIANTE DE LA RELACIÓN DEL ÍNDICE DE  
FERMENTACIÓN DE CACAO CCN-51 DE AROMA CON  
CARACTERÍSTICAS CROMÁTICAS DE LA ALMENDRA DE CACAO**

**AUTOR:**

**WILVER HUMBERTO SANTANA ALVARADO**

**Guayaquil - Ecuador**

**2024**

## RESUMEN

Esta investigación profundiza en el estudio de la influencia de la fermentación sobre las propiedades cromáticas de las almendras de cacao de las variedades EET-103, EET-95, EET-96 e IMC-67. Se aplicó un enfoque cuantitativo mediante técnicas estadísticas avanzadas como el análisis descriptivo y correlacional, el análisis de conglomerados, el análisis de componentes principales (ACP) y modelos lineales generalizados y mixtos, para evaluar cómo las variaciones en el proceso de fermentación influyen en los componentes de color  $L^*$ ,  $a^*$  y  $b^*$  del sistema CIELAB. Es esencial considerar cómo estos hallazgos sobre la influencia significativa de la duración de la fermentación en la luminosidad y coloración de las almendras pueden ser aplicados para mejorar las prácticas de fermentación potenciando así la calidad visual y organoléptica del cacao. También se exploraron las interacciones entre la variedad de cacao y los efectos de la fermentación sobre las características cromáticas, identificando modelos específicos que demuestran cómo la fermentación afecta distintas propiedades cromáticas. Estos hallazgos permiten recomendar prácticas de fermentación óptimas para alcanzar una calidad de cacao deseada y contribuyen al conocimiento científico necesario para estandarizar y optimizar los procesos de producción de chocolate, asegurando la consistencia y calidad del producto final.

**Palabras clave:** fermentación, cacao, propiedades cromáticas, análisis multivariante, CIELAB, modelos lineales generalizados y mixtos, calidad del chocolate.

## ABSTRACT

This research delves into the study of the influence of fermentation on the chromatic properties of cacao beans from the EET-103, EET-95, EET-96 and IMC-67 varieties. A quantitative approach was applied using advanced statistical techniques such as descriptive and correlational analysis, cluster analysis, principal component analysis (PCA) and generalized mixed linear models to evaluate how variations in the fermentation process influence the  $L^*$ ,  $a^*$  and  $b^*$  color components of the CIELAB system. It is essential to consider how these findings on the significant influence of fermentation duration on the brightness and coloration of the beans can be applied to improve fermentation practices, thereby enhancing the visual and organoleptic quality of cacao. Interactions between cacao variety and the effects of fermentation on chromatic characteristics were also explored, identifying specific models that demonstrate how fermentation affects different chromatic properties. These findings allow for the recommendation of optimal fermentation practices to achieve the desired cacao quality and contribute to the scientific knowledge necessary to standardize and optimize chocolate production processes, ensuring the consistency and quality of the final product.

**Keywords:** fermentation, cacao, chromatic properties, multivariate analysis, CIELAB, generalized and mixed linear models, chocolate quality.

## DEDICATORIA

A dos seres maravillosos que con su cariño fueron uno de los pilares fundamentales en mi vida, que están junto a mí y viven en mi corazón esposa e hija, las quiero mucho Michelly Calero y Gisselle Santana.

A la mujer que más admiro por su fuerza y valentía para enfrentarse a este mundo. Olga Alvarado, eres mi madre, mi amiga y el faro que me guía en todo momento y en cada una de las decisiones y etapas de mi vida, gracias por todo lo que haces día a día por mí.

A mis hermanos Ángel y Carlos que a la distancia están pendientes de mí y que luchan día a día por sus familias a los que también admiro y respeto. Mis hermanos queridos.

A mis amigos cercanos Marlon, Rommel, Wilma, Raquel y Anita, gracias por ser mi segunda mi familia, su apoyo incondicional es el impulso que me permite crecer cada día como persona y ser humano.

## **AGRADECIMIENTO**

Agradezco a Dios por ayudarme a llegar hasta este momento tan importante de mi vida que es la graduación, agradezco también a mi madre por darme el apoyo necesario para seguir adelante.

Al Doctor Omar Ruiz, por brindarme la oportunidad de formar parte de su equipo de trabajo.

A la Escuela Superior Politécnica del Litoral y a la Facultad de Ciencias Naturales y Matemáticas, por los conocimientos impartidos a través de todos y cada uno de los profesores que contribuyeron en mi formación profesional.

## **DECLARACIÓN EXPRESA**

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

---

WILVER HUMBERTO SANTANA ALVARADO

# **TRIBUNAL DE GRADUACIÓN**

---

Luis Pilacuan Bonete Ph.D.  
PRESIDENTE

---

Omar Ruiz Barzola Ph.D.  
TUTOR

---

Francisco Moreira Villegas Mgtr.  
EVALUADOR

# TABLA DE CONTENIDO

CAPÍTULO 1 .....	1
1. INTRODUCCIÓN.....	1
1.1. Antecedentes.....	1
1.2. Descripción del problema.....	2
1.3. Objetivos.....	3
1.3.1. General .....	3
1.3.2. Específicos .....	3
1.4. Hipótesis.....	3
1.5. Alcance.....	3
CAPÍTULO 2 .....	5
2. MARCO TEÓRICO .....	5
2.1. Fermentación del Cacao.....	5
2.1.1. Efecto en la Calidad del Cacao.....	5
2.1.2. Impacto en las Características Sensoriales y Físico-Químicas .....	5
2.2. Características cromáticas del cacao.....	6
2.2.1. Importancia de las características cromáticas .....	6
2.2.2. Métodos de evaluación.....	6
2.2.3. Relación entre el color y los compuestos bioactivos .....	6
2.3. Análisis de Datos en Investigación Agrícola.....	6
2.4. Análisis Multivariante y Visualización .....	7
2.5. Modelos Predictivos en Agricultura .....	7
2.6. Modelos estadísticos utilizados en la investigación.....	7
2.6.1. Regresión lineal múltiple.....	7
2.6.2. Modelos Lineales Generalizados y Mixtos.....	8
2.6.3. Análisis de componentes principales (ACP).....	8
2.6.4. Herramientas de Visualización de Datos.....	8



2.6.5.	Análisis de Correlaciones .....	9
2.6.6.	Heatmaps .....	9
CAPÍTULO 3 .....		10
3.	METODOLOGÍA .....	10
3.1.	Descripción de la base de datos .....	10
3.2.	Definición y descripción de las variables.....	11
3.3.	Recolección de datos.....	12
3.3.1.	Método de recolección de datos .....	13
3.3.2.	Representatividad y alcance de la muestra .....	13
3.3.3.	Periodo de estudio .....	13
3.3.4.	Unidades de observación.....	13
3.3.5.	Cantidad de observaciones.....	13
3.4.	Técnica de Investigación y Justificación del Enfoque.....	14
3.4.1.	Enfoque cuantitativo.....	14
3.5.	Proceso de fermentación .....	14
3.6.	Técnicas Estadísticas Aplicadas .....	15
3.6.1.	Herramientas de Visualización de Datos.....	15
3.6.2.	Correlaciones.....	15
3.7.	Análisis Multivariante y visualización .....	16
3.7.1.	Análisis de conglomerados .....	16
3.7.2.	Análisis de Componentes Principales (ACP) y Biplot .....	16
3.8.	Modelos predictivos en agricultura.....	17
3.8.1.	Análisis lineal múltiple .....	17
3.8.2.	Modelo lineal generalizado mixto (MLGM) .....	17
3.9.	Metodología Específica.....	17
3.9.1.	Análisis exploratorio de datos .....	17
3.9.2.	Supuestos de los Modelos de Regresión .....	18
3.9.3.	Modelos de regresión aplicados.....	18

3.9.4.	Modelos lineales generalizados mixtos (MLGM) .....	19
3.10.	Técnicas multivariantes .....	21
3.11.	Herramientas y software utilizado .....	22
3.12.	Procedimiento analítico.....	23
3.12.1.	Flujo de trabajo del análisis de datos.....	23
3.13.	Validación de modelos.....	24
3.13.1.	Validación cruzada de salida única (LOOCV).....	24
3.14.	Selección del mejor modelo.....	25
3.14.1.	Criterios de selección del mejor modelo: AIC Y BIC.....	25
3.15.	Nivel de Significancia de las Pruebas Estadísticas .....	25
CAPÍTULO 4 .....		26
4.	RESULTADOS .....	26
4.1.	Análisis de los Resultados .....	26
4.1.1.	Descriptores estadísticos por variable cromática y variedad de cacao .....	26
4.2.1.	Análisis de la variabilidad de cada variedad de cacao.....	28
4.2.2.	Comparación entre variedades de cacao .....	29
4.2.3.	Diagrama de dispersión luminosidad vs. Fermentación .....	30
4.2.4.	Diagrama de dispersión verde/rojo (RG) vs. Fermentación .....	31
4.2.5.	Diagrama de dispersión luminosidad amarillo/Azul (YB) vs. Fermentación.....	32
4.2.6.	Diagrama de cajas Luminosidad vs. Variedad .....	33
4.2.7.	Diagrama de cajas luminosidad verde/rojo vs. Variedad.....	34
4.2.8.	Diagrama de cajas Luminosidad Amarillo/Azul vs. Variedad.....	35
4.2.9.	Diagrama de cajas índice de fermentación vs. Variedad .....	36
4.3.1.	Visualización de la variabilidad de los datos mediante ACP.....	37
4.3.2.	Interpretación de los componentes principales.....	38
4.4.1.	Análisis de clusters para la segmentación de datos de fermentación del cacao. ....	40
4.4.2.	Mapa de calor de la variabilidad cromática en almendras de cacao bajo diversas condiciones de fermentación.....	43

4.4.3.	Análisis de la relación entre luminosidad y fermentación por variedad .....	44
4.4.4.	Evaluación de la relación entre los componentes de color y la fermentación. ....	45
4.5.1.	Descripción de los modelos .....	48
4.5.2.	Análisis de variabilidad en fermentación por variedad usando modelo lineal mixto.....	51
4.5.3.	Análisis de la normalidad de residuos en modelos estadísticos. ....	52
CAPÍTULO 5 .....		55
5.	CONCLUSIONES Y RECOMENDACIONES .....	55
5.1.	Conclusiones .....	55
5.2.	Recomendaciones .....	56
6.	BIBLIOGRAFIA.....	57
ANEXOS .....		60

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Diagrama de flujo del Proceso de Preparación y Análisis de Datos.....	12
<b>Figura 2.</b> Base de datos de las diferentes variedades de cacao y el índice de fermentación .....	14
<b>Figura 3.</b> Diagrama de flujo del proceso de modelado. ....	19
<b>Figura 4.</b> Muestra un mapa de calor que representa la correlación entre diferentes variables. <b>Error!</b>	
<b>Bookmark not defined.</b>	
<b>Figura 5.</b> Biplot de Componentes Principales.....	22
<b>Figura 6.</b> Análisis de Correlación entre Características Cromáticas y Fermentación de las Almendras de Cacao. ....	29
<b>Figura 7.</b> Diagrama de dispersión Luminosidad vs. Fermentación.....	30
<b>Figura 8.</b> Diagrama de dispersión Luminosidad Verde/Rojo (RG) vs. Fermentación. ....	31
<b>Figura 9.</b> Diagrama de dispersión Luminosidad Amarillo/Azul (YB) vs. Fermentación .....	32
<b>Figura 10.</b> Diagrama de cajas Luminosidad vs. Variedad .....	33
<b>Figura 11.</b> Diagrama de cajas Luminosidad Verde/Rojo vs. Variedad.....	34
<b>Figura 12.</b> Diagrama de cajas Luminosidad Amarillo/Azul vs. Variedad .....	35
<b>Figura 13.</b> Diagrama de cajas índice de fermentación vs. Variedad.....	36
<b>Figura 14.</b> Gráfico de sedimentación .....	37
<b>Figura 15.</b> Análisis Detallado de Componentes Principales para Características Cromáticas del Cacao.....	38
<b>Figura 16.</b> Índice de fermentación de cacao CCN51 y sus características cromáticas. ....	39
<b>Figura 17.</b> Método Partitioning Around Medoids (PAM) para identificar el número óptimo de clústeres.....	40
<b>Figura 18.</b> Biplot de Clústeres formados por la técnica PAM .....	41
<b>Figura 19.</b> Determinación del Número Óptimo de Clústeres mediante el Método K-means .....	42
<b>Figura 20.</b> Biplot de Clústeres Formados por la Técnica PAM .....	42
<b>Figura 21.</b> Mapa de Calor de la Variabilidad Cromática en Almendras de Cacao bajo Diversas Condiciones de Fermentación .....	43
<b>Figura 22.</b> Diagrama de Dispersión de Luminosidad vs. Fermentación por Variedad .....	44
<b>Figura 23.</b> Diagrama de Dispersión de RG vs. Fermentación por Variedad.....	45
<b>Figura 24.</b> Diagrama de Dispersión de YB vs. Fermentación por Variedad.....	47
<b>Figura 25.</b> Resultados del Modelo Lineal Mixto (REML) para la Fermentación del Cacao .....	52

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Resumen del Proceso de Modelado .....	10
<b>Tabla 2.</b> Definición y descripción de las variables.....	11
<b>Tabla 3.</b> Descriptivos para cada variable .....	26
<b>Tabla 4.</b> Descriptivos de cada variable por variedad de la semilla. ....	28
<b>Tabla 5.</b> Tabla de ANOVA para la Intensidad de Luz Amarilla/Azul entre Variedades de Cacao. ....	35
<b>Tabla 6.</b> Variables Cromáticas en Relación con el Índice de Fermentación .....	47
<b>Tabla 7.</b> Evaluación de Modelos Lineales con Efectos Fijos y Aleatorios .....	48
<b>Tabla 8.</b> Comparación de Modelos Lineales Basados en Criterios AIC y BIC .....	49
<b>Tabla 9.</b> Resultados de Pruebas de Normalidad de los Residuos .....	53
<b>Tabla 10.</b> Resultados de las Pruebas de Efectos de la Variedad de Semilla sobre el Índice de Fermentación.....	53

# CAPÍTULO 1

## 1. INTRODUCCIÓN

### 1.1. Antecedentes

En la industria global del cacao, la variedad CCN-51 sobresale por su contribución tanto económica como agronómica, representando el 30% de las exportaciones nacionales en este sector. Es fundamental para la economía del país, esta variedad impulsa el empleo, ingresos y divisas, destacándose aún más por la demanda creciente a nivel mundial, no solo para la elaboración de chocolate sino también valorado en dietas saludables por las propiedades de sus granos (Lainez, 2019).

El cacao, nativo de las selvas húmedas del hemisferio occidental, se cultiva bajo la sombra de árboles frutales, beneficiando la biodiversidad de los agroecosistemas. Distinguido entre el común o Forastero y el fino y de aroma, el último, menos común, pero de calidad superior, encuentra en el país un liderazgo en su producción (Lainez, 2019)

Además de ser un principal exportador de cacao fino y de aroma, el país se ubica en una posición destacada a nivel mundial con la variedad CCN-51, gracias a su resistencia a enfermedades y alta productividad, aspectos clave para la viabilidad económica y agronómica del cultivo en distintas regiones afectadas por patógenos como *Moniliophthora roreri*, causante de la moniliasis (García, 2020).

El cultivo de esta variedad es crucial, soportando aproximadamente 600 mil familias, lo que equivale al 4.3% de la población activa. No obstante, se enfrenta a desafíos como rendimientos promedio bajos y la gestión de residuos agroindustriales, marcando tanto un reto como una oportunidad para innovar y mejorar la sostenibilidad de la producción (García, 2020).

Los desafíos incluyen adversidades climáticas y una dependencia de agroquímicos, destacando la necesidad de adoptar prácticas agrícolas más sostenibles. La variabilidad climática y el impacto del uso intensivo de agroquímicos en la productividad y el ambiente sugieren un cambio urgente hacia el manejo integrado de plagas y enfermedades (Ministerio de Agricultura de Perú, 2012; Torres, 2012; Sánchez Justillo, 2022; Litardo & Arelisa, 2023).

El análisis de la cadena de valor del cacao refleja una industria sostenible en lo económico, social y ambiental, aunque resalta la importancia de diferenciar las variedades mediante sistemas de trazabilidad para asegurar una compensación justa y fomentar la diversidad de cultivos, manteniendo la sostenibilidad de los cultivos de cacao fino y de aroma (Uvillús Herrera, 2023)

Finalmente, la contribución de las exportaciones de CCN-51 a la economía resalta su papel significativo, afectando directamente las exportaciones no petroleras y el PIB. El estudio de su aporte económico entre 2014 y 2018 ilustra fluctuaciones en el sector, indicando áreas para fortalecer la resiliencia y sostenibilidad ante desafíos climáticos y de mercado (Lema Martínez, 2021).

## **1.2. Descripción del problema**

La calidad del cacao es determinante para la industria chocolatera y para el valor que este producto puede alcanzar en el mercado internacional. En particular, la variedad CCN-51, conocida por su robustez y alto rendimiento, presenta desafíos específicos en términos de uniformidad y características de calidad de sus almendras, especialmente cuando se trata de su perfil de aroma, un atributo crítico para su clasificación como cacao de aroma. Uno de los factores cruciales que afectan estas características es el proceso de fermentación, una etapa esencial en la post-cosecha que influye significativamente en el desarrollo del perfil de sabor y aroma del cacao. Sin embargo, la comprensión de cómo el índice de fermentación afecta específicamente las características cromáticas de la almendra de cacao CCN-51, y cómo estas se correlacionan con la calidad percibida del aroma, aún no está completamente desarrollada. (Helliwell, 2015).

El proceso de fermentación es complejo y su eficacia depende de múltiples variables, incluyendo la duración, temperatura, y técnica de fermentación, que juntas influyen en el resultado final del perfil aromático del cacao. La falta de estándares y prácticas de fermentación optimizadas para la variedad CCN-51 de aroma limita la capacidad de los productores para maximizar el potencial de calidad de sus cosechas, impactando directamente en su competitividad y valor de mercado. Además, existe una necesidad crítica de establecer métodos cuantitativos y objetivos para evaluar la fermentación y su impacto en las características cromáticas de las almendras, que puedan servir como indicadores confiables de la calidad del aroma. (Helliwell, 2015)

Este proyecto busca abordar estos vacíos identificando las relaciones multivariantes entre el índice de fermentación y las características cromáticas de las almendras de cacao CCN-51 de aroma. A través de un enfoque analítico detallado se propone estudiar o identificar las relaciones que puedan orientar hacia la mejora de los procesos de fermentación, contribuyendo a mejorar la calidad y homogeneidad del cacao CCN-51 de aroma y, por ende, su aceptación y valor en el mercado internacional. (Helliwell, 2015).

La carencia de investigaciones dedicadas específicamente a la interacción entre la fermentación y las características cromáticas en la variedad CCN-51, junto con la creciente demanda de cacao de alta calidad por parte de la industria y los consumidores, subraya la importancia y urgencia de este estudio. Abordar este problema no solo beneficiará a los productores de cacao CCN-51, mejorando sus prácticas de post-cosecha y la calidad del

producto, sino que también aportará valiosos insights científicos y técnicos al cuerpo de conocimiento existente sobre la fermentación del cacao y su impacto en la calidad final del producto. (Abad, 2019).

### **1.3. Objetivos**

#### **1.3.1. General**

- Estudiar la relación del índice de fermentación de cacao CCN-51 de aroma con características cromáticas de la almendra de cacao, empleando modelos estadísticos multivariantes, para la determinación de su grado de fermentación.

#### **1.3.2. Específicos**

- Realizar la analítica de datos y un análisis relacional del índice de fermentación empleando características cromáticas ( $L^*$ ,  $a^*$  y  $b^*$ ) de almendras de cacao CCN-51, empleando técnicas univariantes y bivariantes.
- Modelar la relación entre los parámetros de estudio a través de técnicas estadísticas multivariantes.
- Validar estadísticamente los modelos obtenidos, identificando de esta manera el modelo más adecuado.

### **1.4. Hipótesis**

El índice de fermentación está significativamente relacionado con las características cromáticas de las almendras de cacao CCN-51, y los cambios en este proceso influyen de manera predecible en las propiedades cromáticas de las almendras, lo cual puede ser un indicador confiable de la calidad aromática del producto final.

### **1.5. Alcance**

El alcance de este proyecto se centra en la implementación de un modelo matemático multivariante existente, aplicado para evaluar y clasificar la fermentación y las características cromáticas de las almendras de cacao CCN-51. Este modelo, validado científicamente, se utilizará para reemplazar los métodos empíricos convencionales, proporcionando una herramienta precisa para medir el índice de fermentación y sus efectos en las propiedades cromáticas del cacao (Pérez L. , 2023).

Este modelo servirá como un recurso invaluable para ingenieros en software y otros investigadores interesados en la aplicación práctica de estos hallazgos. La meta es que, a



partir de este modelo, se desarrollen software y aplicaciones que automatizarán y estandarizarán la evaluación de la calidad del cacao, haciendo que el proceso sea más eficiente, confiable y uniforme a lo largo de la industria. (Calle P & Maldonado F., 2023)

Además, se espera que la adopción de este modelo por parte de la comunidad científica y tecnológica impulse investigaciones adicionales y el desarrollo de herramientas digitales que faciliten la clasificación objetiva del cacao. Esto no solo mejorará la precisión en la determinación de la calidad, sino que también promoverá prácticas más consistentes en el sector, beneficiando a toda la cadena de producción desde los productores hasta los consumidores finales. (Calle P & Maldonado F., 2023)

Finalmente, el proyecto tiene como alcance aumentar la competitividad del cacao ecuatoriano en el mercado global. Al asegurar una calidad superior y más consistente del cacao de aroma, especialmente de la variedad CCN-51, se espera mejorar la percepción y el valor del producto ecuatoriano a nivel internacional, contribuyendo significativamente al prestigio y al éxito económico del sector cacaotero en Ecuador. (Calle P & Maldonado F., 2023)

# CAPÍTULO 2

## 2. MARCO TEÓRICO

### 2.1. Fermentación del Cacao

La fermentación del cacao es un proceso esencial para transformar las semillas recién extraídas en precursores de los ricos sabores y aromas característicos del chocolate. Este proceso biológico, mediante el cual se desarrollan las cualidades sensoriales deseables del cacao, impacta significativamente sus propiedades físico-químicas (Gonzalez Apolo, 2011).

#### 2.1.1. Efecto en la Calidad del Cacao

Esta etapa es decisiva para el desarrollo de las características sensoriales del cacao. Sin ella, las semillas tendrían sabores amargos y astringentes no deseados. La fermentación adecuada no solo mejora el sabor y aroma, sino que también afecta características físico-químicas como la acidez y la formación de compuestos volátiles, crucial para la calidad del chocolate final. (David Salazar, 2016)

Específicamente, para la variedad CCN-51, conocida por su resistencia y productividad, un control cuidadoso del proceso de fermentación es esencial. La investigación de Cabrera et al. (2021) subraya cómo la madurez de las mazorcas incide directamente en la fermentación, afectando tanto la calidad nutricional como sensorial del licor de cacao, lo que resalta la necesidad de ajustar los protocolos de fermentación a las características específicas de cada variedad de cacao. (David Salazar, 2016)

#### 2.1.2. Impacto en las Características Sensoriales y Físico-Químicas

El adecuado manejo de la fermentación permite mejorar la astringencia, sabor, y aroma del cacao, esenciales para la calidad del chocolate. En particular, la fermentación afecta la acidez, la reducción de polifenoles, y la formación de compuestos volátiles, siendo estas alteraciones fundamentales para el producto final y su procesabilidad.

Por lo antes expuesto, la fermentación del cacao, especialmente de la variedad CCN-51, es un equilibrio entre arte y ciencia. El entendimiento profundo de este proceso y la aplicación de técnicas adecuadas son cruciales para maximizar la calidad del cacao, destacando la importancia de investigaciones como las realizadas por Gutiérrez-Paredes et al. (2022) y Cabrera et al. (2021) para la optimización de este proceso en la industria del cacao.

## **2.2. Características cromáticas del cacao**

Las características cromáticas de las almendras de cacao son fundamentales para determinar su calidad y madurez, influenciando profundamente la percepción del consumidor sobre el producto final. Estas características no solo actúan como indicadores visuales del estado de fermentación y procesamiento, sino que también reflejan la presencia de compuestos bioactivos importantes para las propiedades saludables y el sabor del chocolate (Mathias-Rettig, 2014).

### **2.2.1. Importancia de las características cromáticas**

El color de las almendras de cacao es un criterio de calidad crucial dentro de la industria del chocolate. Los cambios en el color durante la fermentación y el secado son indicativos de la correcta transformación de los precursores del sabor y la eliminación de compuestos indeseables (Orbe Chamorro, 2013).

### **2.2.2. Métodos de evaluación**

Los métodos para evaluar las características cromáticas del cacao incluyen la evaluación visual por expertos, la colorimetría, y la espectrofotometría. La colorimetría, utilizando coordenadas de espacio de color como  $L^*$   $a^*$   $b^*$  en el espacio CIELAB, permite una evaluación cuantitativa y objetiva del color. Este enfoque es fundamental para establecer estándares de color y para realizar comparaciones precisas entre lotes de producción (Mathias-Rettig, 2014).

### **2.2.3. Relación entre el color y los compuestos bioactivos**

Los tonos presentes en los alimentos se originan de compuestos mayormente orgánicos, y el cambio de color durante la fermentación del cacao es un indicador de procesos bioquímicos activos. Este cambio es esencial para el proceso de secado y tostado, con los pigmentos de antocianidina en los granos de cacao descomponiéndose por la acción de las glicosidasas durante la fermentación (Afoakwa, 2012). Los compuestos bioactivos en el cacao, incluidos los polifenoles, carotenoides, y fitoesteroles, tienen propiedades antioxidantes, antiinflamatorias, anticancerígenas y antimicrobianas, contribuyendo a los beneficios para la salud del consumo de cacao (Goudar et al., 2023; Prasedya et al., 2021).

## **2.3. Análisis de Datos en Investigación Agrícola**

El análisis de datos juega un papel crítico en la investigación agrícola, especialmente en estudios centrados en la optimización de procesos productivos y la mejora de la calidad de

productos como el cacao. Las herramientas de visualización de datos y el análisis de correlaciones son fundamentales para comprender la complejidad y dinámica de los factores que influyen en la agricultura. (Lema Martínez, 2021).

#### **2.4. Análisis Multivariante y Visualización**

El análisis de conglomerados, el uso de dendrogramas, heatmaps y el Análisis de Componentes Principales (ACP) junto con Biplots constituyen herramientas estadísticas y visuales fundamentales en la investigación agrícola, particularmente útiles para examinar complejas interacciones y variabilidades en cultivos como el cacao. Estas técnicas permiten a los investigadores identificar patrones y agrupaciones naturales en los datos, ofreciendo insights cruciales para la mejora genética, el manejo agronómico y la optimización de procesos post-cosecha. (Calle P & Maldonado F., 2023)

#### **2.5. Modelos Predictivos en Agricultura**

La aplicación de modelos predictivos en la agricultura es esencial para optimizar la producción y calidad de cultivos como el cacao, facilitando la comprensión y predicción de complejas interacciones entre factores ambientales, genéticos y de manejo. (Christian Salas, 2023).

#### **2.6. Modelos estadísticos utilizados en la investigación**

En esta investigación, se emplean varios modelos estadísticos para analizar la relación entre el índice de fermentación del cacao CCN-51 y sus características cromáticas. Los modelos utilizados incluyen la regresión lineal múltiple, los modelos lineales generalizados y mixtos, y el análisis de componentes principales. A continuación, se describen brevemente cada uno de estos modelos y su aplicación en el contexto de la investigación.

##### **2.6.1. Regresión lineal múltiple**

La regresión lineal múltiple es una técnica estadística que permite modelar la relación entre múltiples variables predictoras independientes y una variable dependiente continua. En el contexto del cacao, se utiliza para entender cómo las variaciones en factores como la duración de la fermentación, temperatura y humedad afectan la calidad del cacao. Esta técnica es crucial para optimizar procesos en la producción de cacao (Burgos, 2022).

### 2.6.2. Modelos Lineales Generalizados y Mixtos

El Modelo Lineal Generalizado (MLG) amplía las capacidades de la regresión lineal permitiendo que la variable dependiente siga diversas distribuciones probabilísticas, adaptándose a datos que no cumplen con las condiciones de normalidad. Los modelos mixtos incorporan tanto efectos fijos como aleatorios, permitiendo modelar la variabilidad específica de subgrupos dentro de los datos. Estos modelos son útiles para evaluar la adaptabilidad de clones de cacao en diferentes condiciones ambientales y para analizar datos jerárquicos o con dependencia estructural (**Gallino Cardona, 2018**).

### 2.6.3. Análisis de componentes principales (ACP)

El Análisis de Componentes Principales (ACP) es una técnica que reduce la complejidad de datos correlacionados a componentes no correlacionados, manteniendo la mayor variabilidad. Es útil para simplificar datos de análisis sensorial, químico o genético del cacao, permitiendo interpretar cómo distintas prácticas de fermentación o condiciones ambientales afectan la calidad del cacao (**Kumari, 2018**).

### 2.6.4. Herramientas de Visualización de Datos

Las herramientas de visualización como diagramas de dispersión, cajas y violines proporcionan medios poderosos e intuitivos para explorar y presentar datos. Estas herramientas permiten identificar patrones, tendencias y anomalías dentro de complejos conjuntos de datos, facilitando la comprensión de las interacciones entre variables. Por ejemplo, mediante el uso de diagramas de dispersión y box plots, González Gordon et al. (2019) pudieron evaluar y caracterizar clones de cacao, destacando la importancia de estas herramientas en el análisis agrícola (Cedeño, 2023).

- **Diagramas de Dispersión:** Permiten visualizar la relación entre dos variables cuantitativas, facilitando la identificación de correlaciones lineales o no lineales entre ellas. En el contexto del cacao, son útiles para explorar la relación entre el tiempo de fermentación y cambios específicos en las propiedades cromáticas (Cedeño, 2023).
- **Diagramas de Cajas (Boxplots):** Proporcionan una representación gráfica de la distribución de los datos, destacando la mediana, los cuartiles y los valores atípicos. Son especialmente valiosos para comparar las características de diferentes variedades de cacao en varias condiciones de fermentación (Cedeño, 2023).
- **Diagramas de Violines:** Combinan las propiedades de los boxplots con una estimación de la densidad de probabilidad de los datos, ofreciendo una visión más rica de la

distribución de los datos. Entenderemos la variabilidad en las características de calidad del cacao bajo diferentes procesos de fermentación (Cedeño, 2023).

### **2.6.5. Análisis de Correlaciones**

El análisis de correlaciones es crucial para identificar la fuerza y dirección de las relaciones entre diversas variables. En la investigación sobre cacao, es particularmente importante comprender cómo distintos procesos de fermentación influyen en las características cromáticas y otras propiedades de calidad. González Gordon et al. (2019) y Maldonado Mateus (2023) proporcionan ejemplos claros de cómo el análisis de conglomerados y herramientas estadísticas pueden utilizarse para discernir patrones y correlaciones significativas entre variables, contribuyendo así a la optimización de prácticas de fermentación y mejora de la calidad del cacao (Suárez, 2020).

### **2.6.6. Heatmaps**

Los heatmaps son herramientas de visualización de datos donde los valores se representan como colores, facilitando la identificación de patrones y la intensidad de características específicas en los datos. Son útiles para representar la variabilidad cromática en diferentes variedades de cacao bajo diversas condiciones de fermentación, lo que ayuda a los investigadores a identificar patrones y diferencias significativas en el proceso de fermentación (Cantos et al., 2017).

# CAPÍTULO 3

## 3. METODOLOGÍA

### 3.1. Descripción de la base de datos

A continuación, se presenta un cuadro resumen del proceso de modelado seguido en esta tesis, que incluye desde la recolección de datos hasta la visualización y presentación de los resultados.

*Tabla 1. Resumen del Proceso de Modelado*

<b>Paso</b>	<b>Descripción</b>
<b>Recolección de Datos</b>	<ul style="list-style-type: none"><li>- Método de recolección de datos</li><li>- Representatividad y alcance de la muestra</li><li>- Periodo de estudio</li><li>- Unidades de observación</li><li>- Cantidad de observaciones</li></ul>
<b>Preparación de Datos</b>	<ul style="list-style-type: none"><li>- Limpieza de datos</li><li>- Normalización de datos</li></ul>
<b>Análisis Exploratorio de Datos</b>	<ul style="list-style-type: none"><li>- Estadísticas descriptivas</li><li>- Visualizaciones (histogramas, gráficos de dispersión)</li></ul>
<b>Selección de Modelos</b>	<ul style="list-style-type: none"><li>- Evaluación de supuestos de regresión</li><li>- Selección de modelos adecuados</li></ul>
<b>Ajuste de Modelos Estadísticos</b>	<ul style="list-style-type: none"><li>- Regresión Lineal Múltiple</li><li>- Modelos Lineales Generalizados Mixtos (MLGM)</li></ul>
<b>Validación de Modelos</b>	<ul style="list-style-type: none"><li>- Validación cruzada (LOOCV)</li><li>- Diagnóstico de residuos</li></ul>
<b>Interpretación de Resultados</b>	<ul style="list-style-type: none"><li>- Análisis de resultados obtenidos</li><li>- Implicaciones biológicas y agronómicas</li></ul>
<b>Visualización y Presentación de Resultados</b>	<ul style="list-style-type: none"><li>- Herramientas de visualización de datos</li><li>- Análisis multivariante y biplots</li></ul>

Elaborado por: Autor

### 3.2. Definición y descripción de las variables

*Tabla 2. Definición y descripción de las variables*

<b>Variable</b>	<b>Tipo de Variable</b>	<b>Descripción</b>	<b>Rango/Niveles</b>	<b>Uso en Análisis</b>
Variedad	Categórica	Representa las diferentes variedades de cacao utilizadas en el estudio.	IMC-67, EET-103, EET-96, EET-95	Utilizada como factor en los modelos estadísticos para evaluar diferencias entre variedades.
Luminosidad (L*)	Numérica continua	Medida de la luminosidad de las almendras de cacao, parte del espacio de color CIELAB.	0 a 100	Utilizada como variable predictora en los modelos de regresión para evaluar su efecto en las características del cacao.
Índice Rojo-Verde (a*)	Numérica continua	Medida del índice de color rojo-verde de las almendras de cacao, parte del espacio de color CIELAB.	Valores positivos indican rojo, valores negativos indican verde.	Utilizada como variable predictora en los modelos de regresión.
Índice Amarillo-Azul (b*)	Numérica continua	Medida del índice de color amarillo-azul de las almendras de cacao, parte del espacio de color CIELAB.	Valores positivos indican amarillo, valores negativos indican azul.	Utilizada como variable predictora en los modelos de regresión.
Tiempo de Fermentación	Numérica continua	Número de días que las almendras de cacao fueron sometidas al proceso de fermentación.	1 a 7 días	Utilizada como variable predictora para evaluar su efecto en las propiedades cromáticas del cacao.
Índice de Fermentación	Numérica continua	Medida que representa el grado de fermentación alcanzado por las almendras de cacao.	Porcentaje de fermentación, generalmente entre 0% y 100%.	Variable dependiente en varios modelos para evaluar el impacto de otras variables en el grado de fermentación.

**Elaborado por:** Autor



### 3.3. Recolección de datos

#### Proceso de Preparación y Análisis de Datos

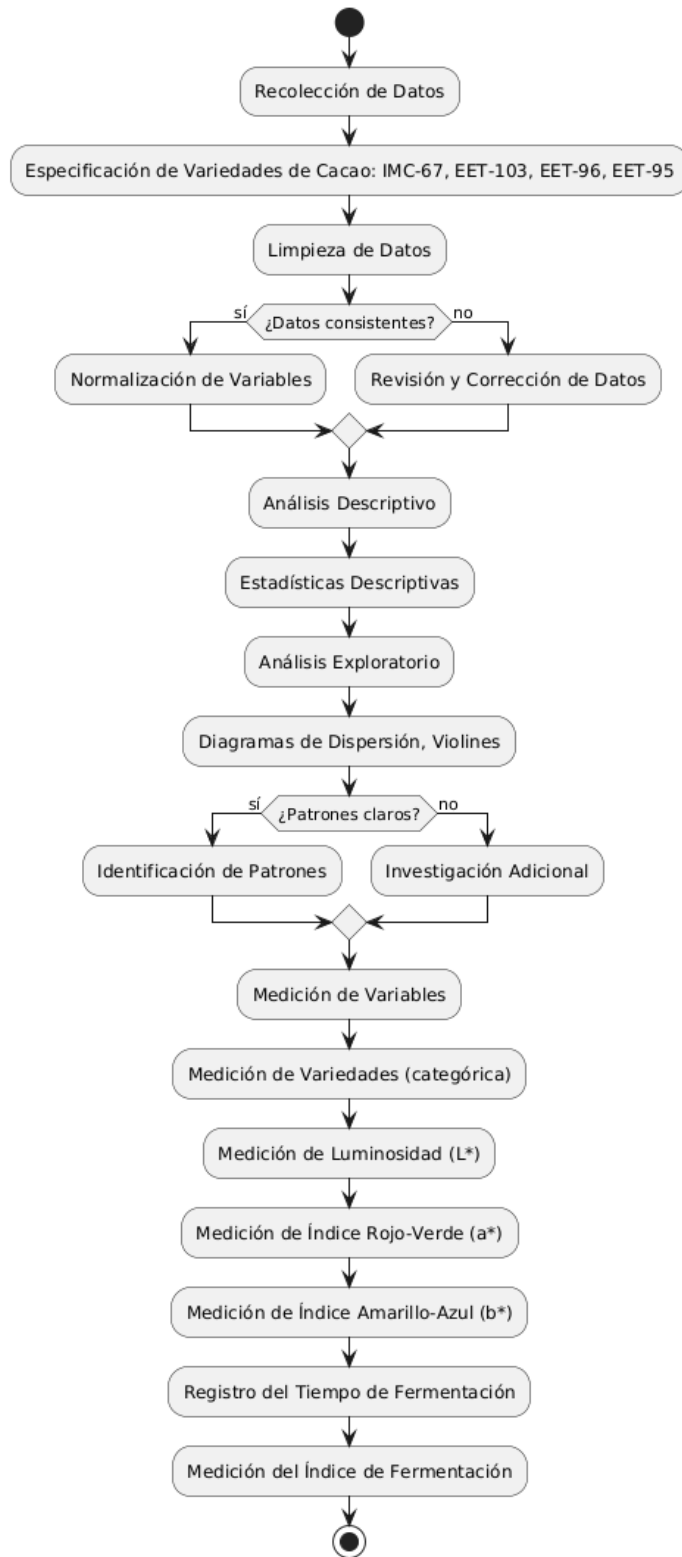


Figura 1. Diagrama de flujo del Proceso de Preparación y Análisis de Datos

Elaborado por: Autor

### **3.3.1. Método de recolección de datos**

La recolección de datos se llevó a cabo bajo estrictos estándares metodológicos para garantizar la precisión y la integridad de la información. Se emplearon técnicas normalizadas y se utilizaron instrumentos de medición calibrados para cuantificar las propiedades cromáticas y otros parámetros físico-químicos relevantes. Los protocolos de fermentación fueron rigurosamente monitoreados para asegurar la fiabilidad en la detección de los cambios bioquímicos que ocurren en las almendras de cacao durante el proceso. (Vasallao, 2019).

### **3.3.2. Representatividad y alcance de la muestra**

El muestreo abarcó datos representativos de diversas fincas ubicadas en la región de Quevedo reconocida por su importancia dentro del sector cacaotero ecuatoriano. La muestra refleja una gama representativa de condiciones agrícolas, incluyendo variaciones climáticas y diferencias en las prácticas de cultivo y fermentación, proporcionando así un marco de análisis aplicable a un contexto amplio y significativo dentro del ámbito agronómico nacional. (Vasallao, 2019).

### **3.3.3. Periodo de estudio**

El periodo cubierto por la recolección de datos corresponde a fases específicas de producción del cacao, documentadas y verificadas a través de la información adjunta al dataset. Estos periodos reflejan momentos clave dentro del ciclo de fermentación y procesamiento post-cosecha (Vasallao, 2017).

### **3.3.4. Unidades de observación**

Las unidades de observación corresponden a mediciones específicas de almendras de cacao, registradas a lo largo de diferentes fases del proceso de fermentación. La metodología aplicada ha permitido analizar la evolución de las propiedades del cacao en momentos críticos, desde el inicio del proceso de fermentación hasta el punto en que las almendras de cacao alcanzan el estado idóneo para su uso final (Vasallao, 2017).

### **3.3.5. Cantidad de observaciones**

El número de observaciones recogidas y analizadas ascendió a 40, repartidas equitativamente entre cuatro variedades de cacao: IMC-67, EET-103, EET-96, y EET-95. Estas variedades se seleccionaron estratégicamente por su relevancia comercial y diferencias potenciales en la fermentación y propiedades finales. La elección de diez muestras por

variedad busca asegurar robustez estadística y confiabilidad en los análisis comparativos, enfocándose en la diversidad genética y su influencia sobre los resultados del proceso de fermentación (Gutierrez-Paredes, 2022).

Se observa en la Figura 2, las variedades de cacao se distribuyen en una matriz de observaciones que reflejan las diferentes etapas de fermentación, lo que permite un análisis detallado de los cambios en las propiedades cromáticas y fisicoquímicas asociados con cada variedad durante este proceso crítico.

Variedad	Réplica	L	a	b	í_f
IMC-67	DOR1	54,61	-7,19	19,86	0,364815
IMC-67	DOR2	54,63	-7,16	19,87	0,395983
IMC-67	DOR3	54,66	-7,15	19,85	0,391152
IMC-67	DOR4	54,67	-7,21	19,87	0,588115
IMC-67	DOR5	54,69	-7,19	19,83	0,586283
IMC-67	DOR6	54,7	-7,19	19,85	0,843866
IMC-67	DOR7	54,73	-7,18	19,84	0,82197
IMC-67	DOR8	54,73	-7,13	19,89	0,984436
IMC-67	DOR9	54,73	-7,18	19,89	0,965649
IMC-67	DOR10	54,75	-7,19	19,93	1,066372
EET-103	DOR1	55,45	-6,76	22,72	1,070175
EET-103	DOR2	55,47	-6,8	22,71	0,398625

*Figura 2. Fragmento del Conjunto de Datos de Variedades de Cacao y Propiedades Cromáticas.*

Elaborado por: Autor

### 3.4. Técnica de Investigación y Justificación del Enfoque

#### 3.4.1. Enfoque cuantitativo

Para este estudio se implementó un enfoque cuantitativo, utilizando modelos estadísticos para analizar datos numéricos que representan características físicas y químicas de las almendras de cacao. Este enfoque permite una evaluación objetiva y precisa de las relaciones entre las variables de estudio, como son los efectos de distintos parámetros de fermentación en las propiedades cromáticas y de calidad del cacao. La metodología cuantitativa es adecuada para este tipo de análisis debido a su capacidad para manejar grandes volúmenes de datos y proporcionar resultados verificables y replicables, fundamentales para la toma de decisiones basada en evidencia en el ámbito agrícola (Gutierrez, 2011).

### 3.5. Proceso de fermentación

Comienza con la recolección de las vainas de cacao, seguida de la extracción de las semillas y su pulpa. Las semillas se colocan en contenedores adecuados donde fermentan por varios días. Durante este tiempo, la pulpa se descompone gracias a la acción de microorganismos naturales, iniciando con levaduras y seguido por bacterias lácticas y acéticas, pasando por etapas de iniciación, acidificación, y secado (Gonzalez Apolo, 2011).

Gutierrez-Paredes et al. (2022) destacan la importancia de seleccionar adecuadamente la tecnología de fermentación para mejorar la calidad sensorial del cacao. En su estudio, demuestran cómo la fermentación en cajón de madera, comparada con otros métodos, optimiza el proceso fermentativo del cacao CCN-51, logrando así un mayor grado de fermentación y, por ende, una mejor calidad del producto final.

### 3.6. Técnicas Estadísticas Aplicadas

En esta sección se describen las técnicas estadísticas aplicadas específicamente en el análisis de los datos recolectados sobre el cacao.

#### 3.6.1. Herramientas de Visualización de Datos

Para explorar y presentar los datos de las propiedades cromáticas del cacao, se utilizaron diversas herramientas de visualización:

- **Diagramas de Dispersión:** Se utilizaron para visualizar la relación entre el tiempo de fermentación y los cambios específicos en las propiedades cromáticas ( $L^*$ ,  $a^*$ ,  $b^*$ ). En nuestro estudio, los diagramas de dispersión permitieron identificar la correlación entre el tiempo de fermentación y la luminosidad de las almendras de cacao, destacando cómo periodos de fermentación más largos influenciaban negativamente la claridad del cacao.
- **Diagramas de Cajas (Boxplots):** Estos fueron empleados para comparar la distribución de las propiedades cromáticas entre las diferentes variedades de cacao en varias condiciones de fermentación. En la investigación, los boxplots facilitaron la comparación entre las variedades IMC-67, EET-103, EET-96, y EET-95 bajo diferentes tiempos de fermentación, resaltando las diferencias y similitudes en las propiedades cromáticas.
- **Diagramas de Violines:** Combinando las propiedades de los boxplots con estimaciones de densidad, estos diagramas proporcionaron una visión más detallada de la distribución de las propiedades cromáticas bajo diferentes tiempos de fermentación. Los diagramas de violines fueron esenciales para visualizar la variabilidad y la distribución de las medidas de color ( $L^*$ ,  $a^*$ ,  $b^*$ ) en función del tiempo de fermentación, ayudando a identificar patrones específicos en la distribución de los datos.

#### 3.6.2. Correlaciones

El análisis de correlaciones se llevó a cabo para entender cómo los distintos procesos de fermentación influyen en las propiedades cromáticas del cacao:

- **Análisis de Correlaciones:** Se calcularon coeficientes de correlación de Pearson para evaluar la fuerza y dirección de la relación entre el tiempo de fermentación y las

propiedades cromáticas ( $L^*$ ,  $a^*$ ,  $b^*$ ). En este estudio, se encontraron correlaciones significativas entre el tiempo de fermentación y el índice rojo-verde ( $a^*$ ), así como entre el tiempo de fermentación y el índice amarillo-azul ( $b^*$ ). Estas correlaciones fueron cruciales para determinar el impacto de la duración de la fermentación en la calidad del cacao, proporcionando información valiosa para optimizar el proceso de fermentación.

### 3.7. Análisis Multivariante y visualización

#### 3.7.1. Análisis de conglomerados

En este estudio, utilizamos diversas técnicas para agrupar y visualizar la similitud entre diferentes variedades de cacao.

**Dendrograma:** Empleamos un dendrograma para visualizar la similitud entre las variedades de cacao, mostrando las conexiones de grupos. Esta herramienta fue crucial para entender la estructura de los datos y determinar el número óptimo de conglomerados, permitiéndonos identificar patrones y diferencias significativas entre las variedades de cacao analizadas (Freire L. M., 2017).

- **Heatmap:** Utilizamos un heatmap para visualizar la intensidad de características específicas en diferentes variedades de cacao. Este método facilitó la identificación de patrones en los datos recolectados, permitiéndonos observar cómo las distintas condiciones de fermentación afectan las propiedades cromáticas de las variedades de cacao (Cantos, 2017).

#### 3.7.2. Análisis de Componentes Principales (ACP) y Biplot

**ACP:** Para reducir la complejidad de los datos correlacionados, aplicamos el Análisis de Componentes Principales (ACP). Esta técnica nos permitió transformar los datos a componentes no correlacionados, manteniendo la mayor variabilidad posible. Esto fue útil para simplificar los datos de análisis sensorial, químico y genético del cacao, facilitando la interpretación de los resultados.

- **Biplot:** Utilizamos un Biplot para mostrar las puntuaciones de las muestras en componentes principales y las variables originales. Esta herramienta nos ayudó a interpretar cómo distintas prácticas de fermentación o condiciones ambientales afectan la calidad del cacao, revelando la contribución de cada variable a la variación observada en los datos (Kumari, 2018).

### **3.8. Modelos predictivos en agricultura**

#### **3.8.1. Análisis lineal múltiple**

En nuestro estudio, aplicamos la técnica de regresión lineal múltiple para modelar la relación entre varias variables predictoras independientes (como la duración de la fermentación, temperatura y humedad) y la variable dependiente continua (calidad del cacao). Esta herramienta nos permitió entender mejor cómo las variaciones en estos factores afectan la calidad del cacao.

Utilizando esta técnica, logramos optimizar ciertos procesos, como la cinética de secado del cacao, lo que contribuyó a reducir pérdidas y mejorar la eficiencia de producción. Los resultados obtenidos son consistentes con estudios previos, como los de Burgos, Menéndez, & Bedón (2022) y Nivela Andrade (2020), quienes también aplicaron la regresión lineal múltiple para mejorar diversos aspectos de la producción de cacao.

#### **3.8.2. Modelo lineal generalizado mixto (MLGM)**

En nuestro estudio, utilizamos el Modelo Lineal Generalizado Mixto (MLGM) para ampliar las capacidades de la regresión lineal, permitiendo que la variable dependiente siguiera diversas distribuciones probabilísticas. Esto fue particularmente útil para modelar datos del cacao que no cumplen con las condiciones de normalidad, como la proporción de almendras germinadas o la intensidad de las características sensoriales.

El MLGM nos permitió evaluar la adaptabilidad de diferentes clones de cacao en diversas condiciones ambientales. Este modelo fue eficaz para identificar cómo factores específicos afectan la calidad del cacao y para implementar mejoras en los procesos agrícolas. Gracias al MLGM, logramos seleccionar variedades con mejores características agronómicas y mayor resistencia, alineándose con los resultados de Gallino Cardona (2018) y Cruz & Mullo (2020).

### **3.9. Metodología Específica**

#### **3.9.1. Análisis exploratorio de datos**

El análisis exploratorio de datos se realizó mediante el uso intensivo de estadísticas descriptivas para obtener medias, medianas, modas, rangos, desviaciones estándar y coeficientes de variación, permitiendo una primera comprensión de la centralidad y dispersión de los datos. (Quiroz, 2023).

El análisis exploratorio también se centrará en identificar la presencia de valores atípicos que podrían influir de manera significativa en los resultados de una muestra pequeña. Además, se analizaron las distribuciones de frecuencia y se implementaron visualizaciones gráficas, como histogramas y gráficos de dispersión.

### 3.9.2. Supuestos de los Modelos de Regresión

Para garantizar la precisión y validez de los modelos de regresión aplicados, se verificaron los siguientes supuestos:

- **Linealidad:** Se comprobó mediante la inspección de gráficos de dispersión entre las variables independientes y la variable dependiente.
- **Independencia de los Errores:** Utilizamos la prueba de Durbin-Watson para detectar la autocorrelación entre los residuos.
- **Homocedasticidad:** Se examinó a través de gráficos de residuos vs. valores ajustados para verificar la consistencia en la varianza de los residuos. Además, se aplicaron pruebas estadísticas como la prueba de Breusch-Pagan y la prueba de White para evaluar la homocedasticidad (Gámez-Villazana, 2022).

### 3.9.3. Modelos de regresión aplicados

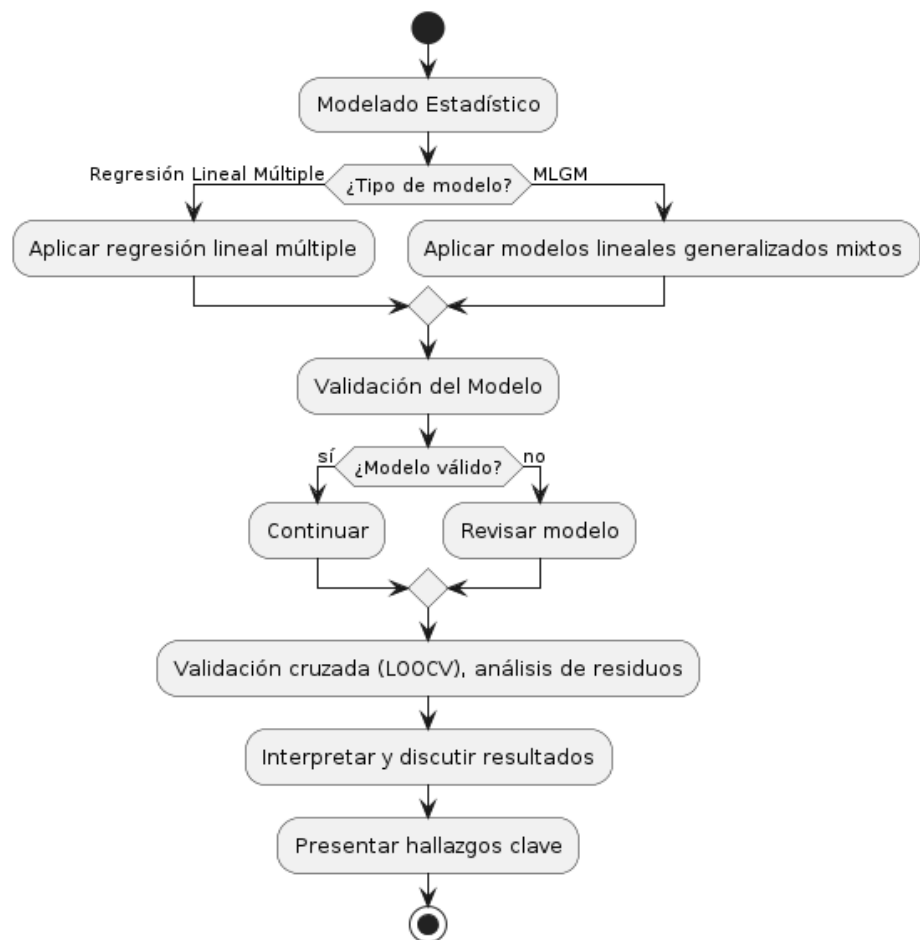
La metodología incluyó varios modelos de regresión para evaluar la influencia de distintas variables predictoras sobre la variable de respuesta. La variable de "Variedad" se trató como un efecto aleatorio en algunos modelos, particularmente en los modelos lineales generalizados mixtos (MLGM), para manejar la variabilidad dentro de los grupos o clústeres de datos (Zuluaga, 2019).

- **Modelo Lineal Simple:** Analizó la relación directa entre variables individuales y la respuesta.
- **Modelo Lineal Múltiple:** Integró múltiples predictores para capturar relaciones más complejas entre las variables, permitiendo una comprensión más profunda de cómo diferentes factores interactúan para afectar la variable dependiente. Previamente se analizó la multicolinealidad de las variables descriptoras a través del Factor de Inflación de la Varianza (VIF) (Ruiz, 2020).
- **Modelo Lineal Generalizado Mixto:** Se empleó para datos que no seguían una distribución normal, ajustando funciones de enlace apropiadas y considerando tanto efectos fijos como aleatorios para manejar la variabilidad dentro de grupos o clusters de datos (Freire O. M., 2019).

### 3.9.4. Modelos lineales generalizados mixtos (MLGM)

**Introducción a los Modelos de Efectos Mixtos:** Para evaluar el efecto de diferentes variables en la fermentación del cacao, se ajustaron varios Modelos Lineales Generalizados Mixtos (MLGM). Los MLGM permiten incluir tanto efectos fijos como aleatorios en el modelo, lo que es útil para analizar datos con estructuras jerárquicas o agrupadas.

#### Proceso de Modelado Estadístico y Validación del Modelo



*Figura 3. Diagrama de flujo del proceso de modelado.*

**Elaborado por:** Autor

- **Efectos Fijos y Aleatorios:** Los efectos fijos ( $\beta$ ) y aleatorios ( $\mu$ ) se diferencian en que los primeros se asocian con niveles específicos de las variables categóricas y los segundos con las variaciones aleatorias entre grupos o sujetos. Los factores adicionales ( $\gamma$ ) aportan variabilidad y pueden ser cruzados o anidados dependiendo del diseño del experimento (Briones, 2020).



## **Modelos utilizados:**

### **1.- Modelo con intercepto aleatorio para la variedad:**

$$Y_{si} = \beta_0 + (1|variedad) + \epsilon_{si}$$

Este modelo básico considera que solo el intercepto varía entre las diferentes variedades de cacao.

### **2.- Modelo con luminosidad como variable fija e intercepto aleatorio para la variedad:**

$$Y_{si} = \beta_0 + \beta_1 \cdot Luminosidad + (1|variedad) + \epsilon_{si}$$

Aquí se evalúa el efecto de la luminosidad en la fermentación, permitiendo que el intercepto varíe entre las variedades.

### **3.- Modelo con RG como variable fija e intercepto aleatorio para la variedad:**

$$Y_{si} = \beta_0 + \beta_1 \cdot RG + (1|variedad) + \epsilon_{si}$$

Este modelo incluye RG (índice rojo-verde) como variable fija, con un intercepto aleatorio para la variedad.

### **4.- Modelo con YB como variable fija e intercepto aleatorio para la variedad:**

$$Y_{si} = \beta_0 + \beta_1 \cdot YB + (1|variedad) + \epsilon_{si}$$

Para capturar la interacción entre diferentes factores, se utilizó un modelo que incluyó efectos aleatorios de las variables de interés, como variedad de cacao e índice de fermentación.

### **5.- Modelo con luminosidad y RG como variables fijas e intercepto aleatorio para la variedad:**

$$Y_{si} = \beta_0 + \beta_1 \cdot Luminosidad + \beta_2 \cdot RG + (1|variedad) + \epsilon_{si}$$

Este modelo incluye las dos variables principales de interés (Luminosidad y RG) para evaluar su efecto combinado en la fermentación.

### **6.- Modelo con luminosidad, RG y YB como variables fijas e intercepto aleatorio para la variedad:**

$$Y_{si} = \beta_0 + \beta_1 \cdot Luminosidad + \beta_2 \cdot RG + \beta_3 \cdot YB + (1|variedad) + \epsilon_{si}$$

Se considera el efecto combinado de todas las variables cromáticas principales, permitiendo variabilidad en el intercepto.

### **3.10. Técnicas multivariantes**

El uso de técnicas multivariantes fue esencial para analizar las características cromáticas y otros atributos de las variedades de cacao. Estas técnicas permitieron visualizar y modelar la relación entre múltiples variables de forma efectiva.

#### **3.10.1. Análisis de Componentes Principales (ACP) y Biplot**

En nuestro estudio, aplicamos el Análisis de Componentes Principales (ACP) para reducir la complejidad de los datos correlacionados a componentes no correlacionados, manteniendo la mayor variabilidad posible. Esto nos permitió simplificar los datos de análisis sensorial, químico y genético del cacao, facilitando la interpretación de los resultados. Utilizamos un Biplot para mostrar las puntuaciones de las muestras en componentes principales y las variables originales, ayudándonos a interpretar cómo distintas prácticas de fermentación o condiciones ambientales afectan la calidad del cacao, revelando la contribución de cada variable a la variación observada en los datos (Kumari, 2018).

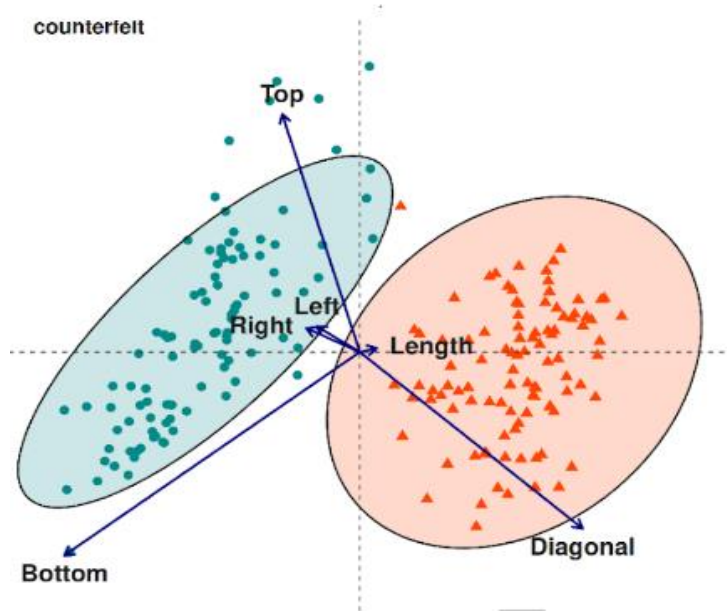
#### **3.10.2. Análisis de componentes principales (ACP) y biplots**

- **ACP:** Se utilizó para reducir la dimensionalidad del conjunto de datos, transformando variables correlacionadas en un número menor de variables no correlacionadas llamadas componentes principales. La matriz de covarianza se descompuso como sigue:

$$X=U\Sigma V^T$$

Donde  $X$  es la matriz de datos original,  $U$  y  $V$  son matrices ortogonales de vectores propios, y  $\Sigma$  es una matriz diagonal de valores propios.

- **Biplots:** Se emplearon para visualizar los resultados del ACP, mostrando tanto las observaciones como las variables en un mismo gráfico bidimensional, lo que ayuda a interpretar las relaciones entre las variables y las observaciones. (Cedeño, 2023).



*Figura 4. Biplot de Componentes Principales*

**Fuente:** Agron Info Tech (2020)

### 3.11. Herramientas y software utilizado

**Software Estadístico:** Los análisis se realizaron utilizando software estadístico como R y Python, empleando librerías específicas para análisis estadístico y visualización de datos. (Gonzalez Apolo, 2011).

### 3.12. Procedimiento analítico

#### 3.12.1. Flujo de trabajo del análisis de datos

El flujo de trabajo del análisis de datos se diseñó meticulosamente para garantizar la precisión y aplicabilidad de los resultados obtenidos en la investigación sobre el cacao. Se estructuró en las siguientes etapas:

**Preparación de Datos:** Antes de proceder con análisis complejos, los datos se limpiaron y prepararon mediante procedimientos estándar:

- **Normalización:** Se normalizaron las características cuantitativas para tener una media de cero y una desviación estándar de uno, usando la fórmula:

$$x_{norm} = \frac{x - \mu}{\sigma}$$

Donde  $x$  es el valor original,  $\mu$  es la media de la variable, y  $\sigma$  es la desviación estándar.

**Selección de Modelos:** La selección de los modelos adecuados dependió del análisis exploratorio y de las características específicas del conjunto de datos:

- **Modelos Estadísticos:** Se eligieron modelos basados en la distribución y la naturaleza de las variables. Para datos continuos y normalmente distribuidos, se optó por modelos lineales.

**Pruebas de Hipótesis:** Para verificar la validez de las relaciones modeladas, se realizaron pruebas estadísticas como:

- **Prueba T:** Utilizada para comparar las medias de dos grupos, se calcula como:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \cdot \sqrt{\frac{2}{n}}}$$

Donde  $\bar{x}_1$  y  $\bar{x}_2$  son las medias de los grupos,  $s_p$  es la desviación estándar agrupada, y  $n$  es el tamaño de cada grupo.

**Interpretación de resultados:** Se discutieron las implicaciones de los resultados estadísticos, concentrándose en cómo estos reflejan los procesos biológicos y agronómicos del cultivo del cacao (Gutierrez, 2011).

### 3.13. Validación de modelos

Para confirmar la validez y robustez de los modelos estadísticos, se implementaron diversas técnicas de validación adaptadas a la naturaleza y al tamaño de la muestra de los datos.

#### 3.13.1. Validación cruzada de salida única (LOOCV)

Dado que la muestra es pequeña, la validación cruzada de salida única se utilizó para evaluar la capacidad predictiva del modelo. En el LOOCV, si tenemos  $n$  observaciones, el modelo se entrena  $n$  veces, cada vez excluyendo una observación diferente del conjunto de entrenamiento y utilizando esa observación como la prueba. La validación cruzada de salida única se calcula como:

$$LOOCV = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Donde  $y_i$  es la observación real excluida del conjunto de entrenamiento  $\hat{y}_i$  es la predicción del modelo sin la observación  $i$ .

- **Prueba de Shapiro-Wilk para Normalidad:** Se aplicó esta prueba para evaluar la normalidad de los residuos. La hipótesis nula que la distribución es normal se prueba como:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \tilde{x}_i)^2}$$

- **Análisis de Influencia:** Se calcularon medidas como la distancia de Cook para identificar posibles puntos de influencia que pudieran sesgar el modelo:

$$D_i = \frac{\left(\frac{r_i^2}{p}\right)}{MSE} \cdot \frac{h_{ii}}{(1-h_i)^2}$$

Donde  $r_i$  es el residual para la observación  $i$ ,  $p$  es el número de parámetros en el modelo, MSE es el error cuadrático medio, y  $h_{ii}$  es el apalancamiento de la observación.

### **3.14. Selección del mejor modelo**

#### **3.14.1. Criterios de selección del mejor modelo: AIC Y BIC**

Para evaluar y seleccionar el mejor modelo, se utilizaron los criterios de Información de Akaike (AIC) y de Información Bayesiana (BIC).

- **AIC (Criterio de Información de Akaike):** El AIC mide la calidad del modelo y penaliza la complejidad del modelo. Un menor valor de AIC indica un mejor ajuste del modelo a los datos, comparado con modelos más complejos.
- **BIC (Criterio de Información Bayesiana):** Similar al AIC, el BIC incluye una penalización más fuerte por el número de parámetros. Un menor valor de BIC indica un mejor ajuste del modelo, reforzando la simplicidad y efectividad del modelo.

Ambos criterios se utilizaron para seleccionar el modelo más adecuado para los datos de fermentación del cacao, considerando la variabilidad cromática bajo diferentes condiciones de fermentación. (Gonzalez Apolo, 2011)

### **3.15. Nivel de Significancia de las Pruebas Estadísticas**

En el presente estudio, todas las pruebas estadísticas se realizaron bajo un nivel de significancia del 5%. Este criterio se estableció para asegurar que los resultados obtenidos sean robustos y reflejen con precisión las relaciones significativas entre las variables cromáticas y el índice de fermentación del cacao CCN-51 de aroma. Al utilizar este umbral, se garantiza que las conclusiones derivadas de los análisis multivariantes y modelos estadísticos aplicados en esta investigación son estadísticamente válidas y confiables.

# CAPÍTULO 4

## 4. RESULTADOS

### 4.1. Análisis de los Resultados

#### 4.1.1. Descriptores estadísticos por variable cromática y variedad de cacao

Se presentan las estadísticas descriptivas para cada variable cromática ( $L^*$ ,  $a^*$ ,  $b^*$ ) y las variedades de cacao (IMC-67, EET-103, EET-96, EET-95).

*Tabla 3. Descriptivos para cada variable*

	Luminosidad	RG	YB	Fermentación
<b>Mínimo</b>	53.57	-7.21	19.83	0.14
<b>Promedio</b>	55.07	-6.84	21.88	0.71
<b>Mediana</b>	55.10	-6.87	22.29	0.79
<b>Coefficiente de Asimetría</b>	-0.01	0.30	-1.07	-0.21
<b>Máximo</b>	56.60	-6.41	22.73	1.09
<b>Des. Estándar</b>	1.10	0.27	1.19	0.27

Elaborado por: Autor

**Luminosidad:** Toma valores en el intervalo 53.57 a 56.60, con un promedio de 55.07 y mediana de 55.10.

Dado que el promedio y la mediana son muy similares el coeficiente de asimetría es relativamente bajo (-0.01), por lo que se puede inferir que el histograma de la variable no presenta sesgo.

**RG:** Toma valores en el intervalo -7.21 a -6.41 con un promedio de -6.84 y mediana de -6.87.

Dado que el promedio y la mediana son muy similares el coeficiente de asimetría es relativamente bajo (0.3), por lo que se puede inferir que el histograma de la variable no presenta sesgo.

**YB:** Toma valores en el intervalo 19.83 a 22.73 con un promedio de 21.88 y mediana de 22.29.

El coeficiente de asimetría es -1.07, lo que indica que la variable podría tener algunos valores atípicos ubicados en los valores más bajos de la variable, sesgándola hacia la izquierda.

**Fermentación:** Toma valores en el intervalo 0.14 a 1.09 con un promedio de 0.71 y mediana de 0.79.

Dado que el promedio y la mediana son muy similares el coeficiente de asimetría es relativamente bajo (-0.21), por lo que se puede inferir que el histograma de la variable no presenta sesgo.

En general todas las variables tienen una dispersión baja, pero las variables YB y Luminosidad tienen las desviaciones estándares más altas dentro del conjunto analizado.

Las variaciones en las variables YB y Luminosidad entre las diferentes variedades de cacao indican que el factor 'Variedad' es determinante en estas características cromáticas. Desde un punto de vista agronómico, esto sugiere que ciertas variedades son genéticamente más propensas a variar en color y luminosidad, lo cual es fundamental para la selección y manejo agronómico de las variedades, buscando uniformidad y estabilidad en las propiedades visuales del cacao.



## 4.2. Análisis de variabilidad por variedad

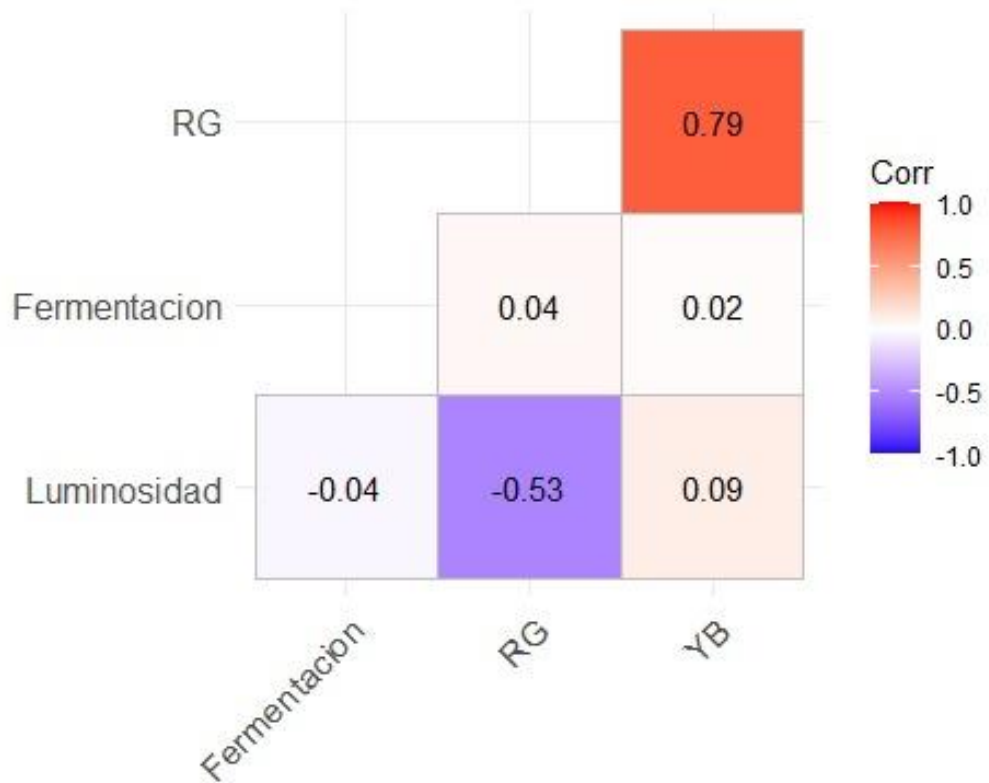
### 4.2.1. Análisis de la variabilidad de cada variedad de cacao

*Tabla 4. Descriptivos de cada variable por variedad de la semilla.*

		EET-103	EET-95	EET-96	IMC-67
Luminosidad	Mínimo	55.45	53.57	56.47	54.61
	Promedio	55.47	53.59	56.55	54.69
	Mediana	55.47	53.59	56.57	54.70
	Coefficiente de Asimetría	0.63	-1.61	-0.39	-0.40
	Máximo	55.49	53.60	56.60	54.75
	Des. Estándar	0.01	0.01	0.05	0.05
RG	Mínimo	-6.80	-6.47	-7.01	-7.21
	Promedio	-6.78	-6.45	-6.96	-7.18
	Mediana	-6.80	-6.47	-6.96	-7.19
	Coefficiente de Asimetría	0.41	0.63	-0.83	0.75
	Máximo	-6.76	-6.41	-6.93	-7.13
	Des. Estándar	0.02	0.03	0.02	0.02
YB	Mínimo	22.28	22.63	22.25	19.83
	Promedio	22.67	22.70	22.27	19.87
	Mediana	22.72	22.72	22.27	19.87
	Coefficiente de Asimetría	-2.63	-1.90	0.16	0.79
	Máximo	22.73	22.72	22.29	19.93
	Des. Estándar	0.14	0.03	0.02	0.03
Índice de Fermentación	Mínimo	0.39	0.40	0.14	0.36
	Promedio	0.73	0.71	0.68	0.70
	Mediana	0.86	0.79	0.70	0.71
	Coefficiente de Asimetría	-0.27	-0.22	-0.16	-0.02
	Máximo	1.07	0.99	1.09	1.07
	Des. Estándar	0.28	0.25	0.33	0.27

*Elaborado por: Autor*

#### 4.2.2. Comparación entre variedades de cacao



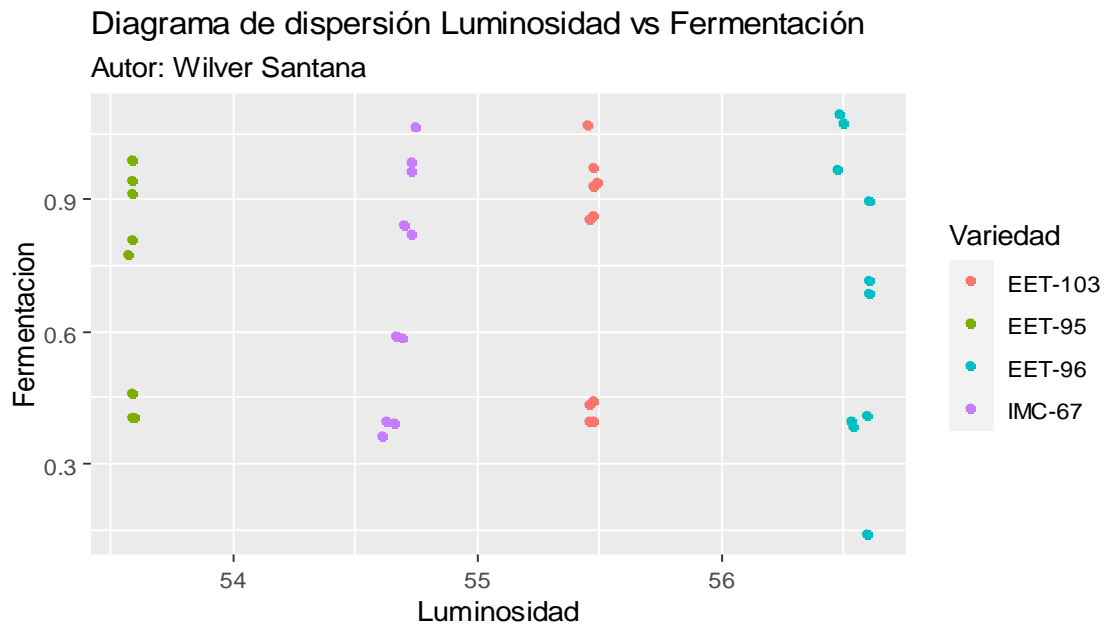
**Figura 5.** Análisis de Correlación entre Características Cromáticas y Fermentación de las Almendras de Cacao.

**Elaborado por:** Autor

Las variables YB y RG, tienen un coeficiente de correlación 0.79 lo que indicaría la relación lineal entre las mismas.

La variable fermentación no muestra relación lineal con ninguna de las variables analizadas, por lo que se buscaran otro tipo de relaciones existentes.

### 4.2.3. Diagrama de dispersión luminosidad vs. Fermentación

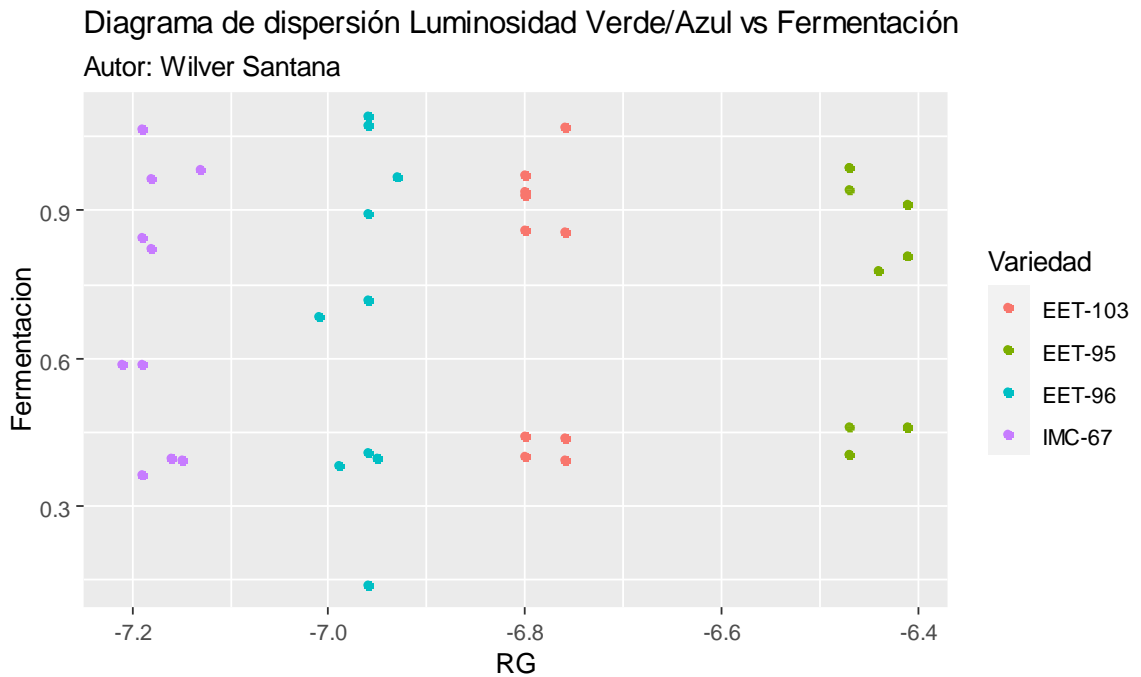


**Figura 6.** Diagrama de dispersión Luminosidad vs. Fermentación

**Elaborado por:** Autor

En el diagrama de dispersión se puede observar que la Variedad EET-96 tiene los índices más altos de luminosidad, seguido de EET-103, IMC-67 y EET-95 quien tiene la luminosidad más baja. No obstante, no existe una relación lineal ya que aumentar la luminosidad no necesariamente aumentará el índice de fermentación.

#### 4.2.4. Diagrama de dispersión verde/rojo (RG) vs. Fermentación



**Figura 7.** Diagrama de dispersión Luminosidad Verde/Rojo (RG) vs. Fermentación.

**Elaborado por:** Autor

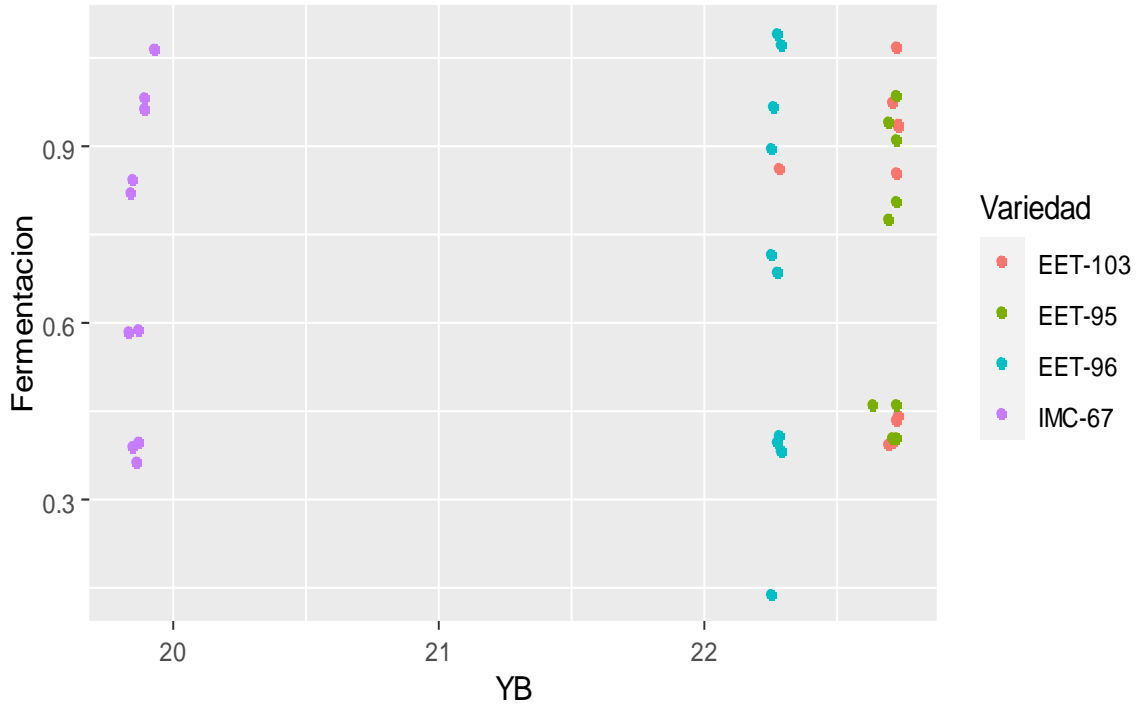
En el diagrama de dispersión se puede observar que la Variedad EET95 tiene los índices más altos de intensidad de luz Roja y Verde, seguida de EET-103, EET-96 y IMC-67 quien tiene la intensidad de color más baja.

No obstante, no existe una relación lineal ya que aumentar la intensidad de la luz Roja/Verde no necesariamente aumentará el índice de fermentación.

#### 4.2.5. Diagrama de dispersión luminosidad amarillo/Azul (YB) vs. Fermentación

Diagrama de dispersión Luminosidad Amarillo/Azul vs Fermentación

Autor: Wilver Santana



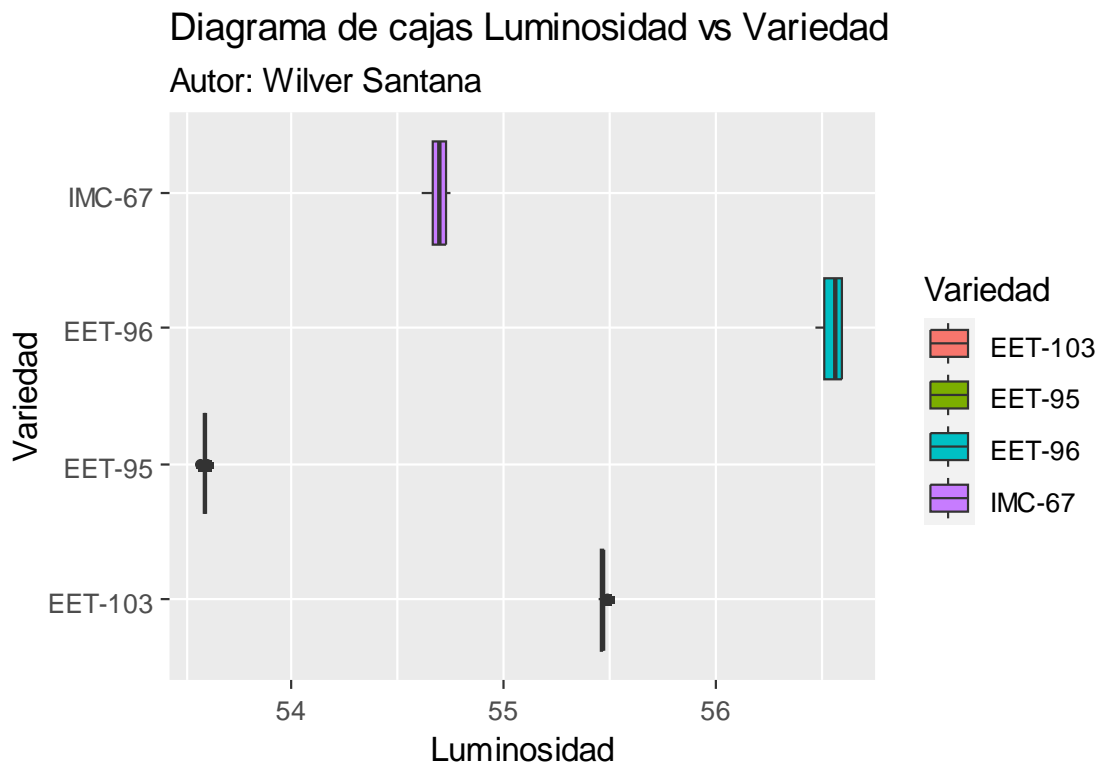
**Figura 8.** Diagrama de dispersión Luminosidad Amarillo/Azul (YB) vs. Fermentación

**Elaborado por:** Autor

En el diagrama de dispersión puede observarse que el comportamiento de las especies EET-103 y EET-95 obtienen índices de Fermentación similares con iguales valores de intensidad de luz Amarilla/Azul.

Seguida por la variedad de semilla EET-96 y por la variedad IMC-67 que obtuvo la menor intensidad de luz Amarilla/Azul.

#### 4.2.6. Diagrama de cajas Luminosidad vs. Variedad



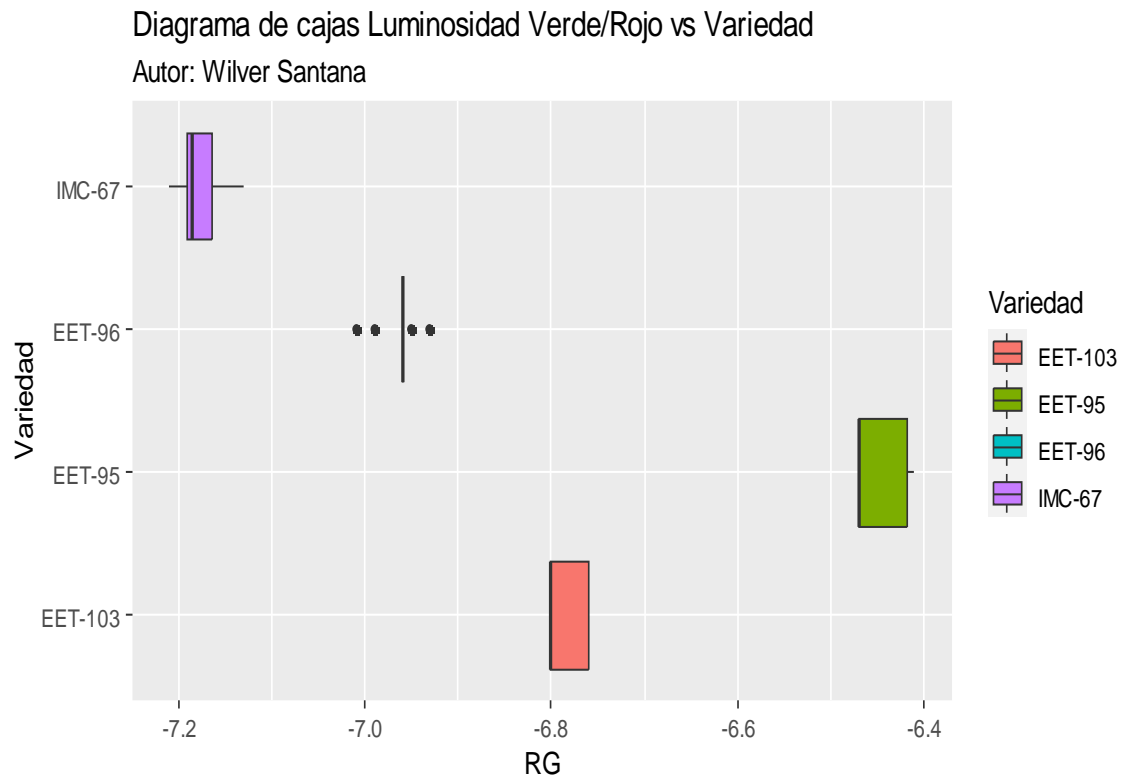
*Figura 9. Diagrama de cajas Luminosidad vs. Variedad*

**Elaborado por:** Autor

Las variedades EET-95 y EET-103 obtuvieron valores de luminosidad de baja dispersión, mientras que la variedad EET-96 tiene los valores “más dispersos” de las cuatro especies analizadas.

Por otro lado, la variedad EET-96 obtuvo los valores de luminosidad más alto, seguido por las variedades EET-103, IMC-67 y EET-95 que obtuvo la luminosidad más baja.

#### 4.2.7. Diagrama de cajas luminosidad verde/rojo vs. Variedad



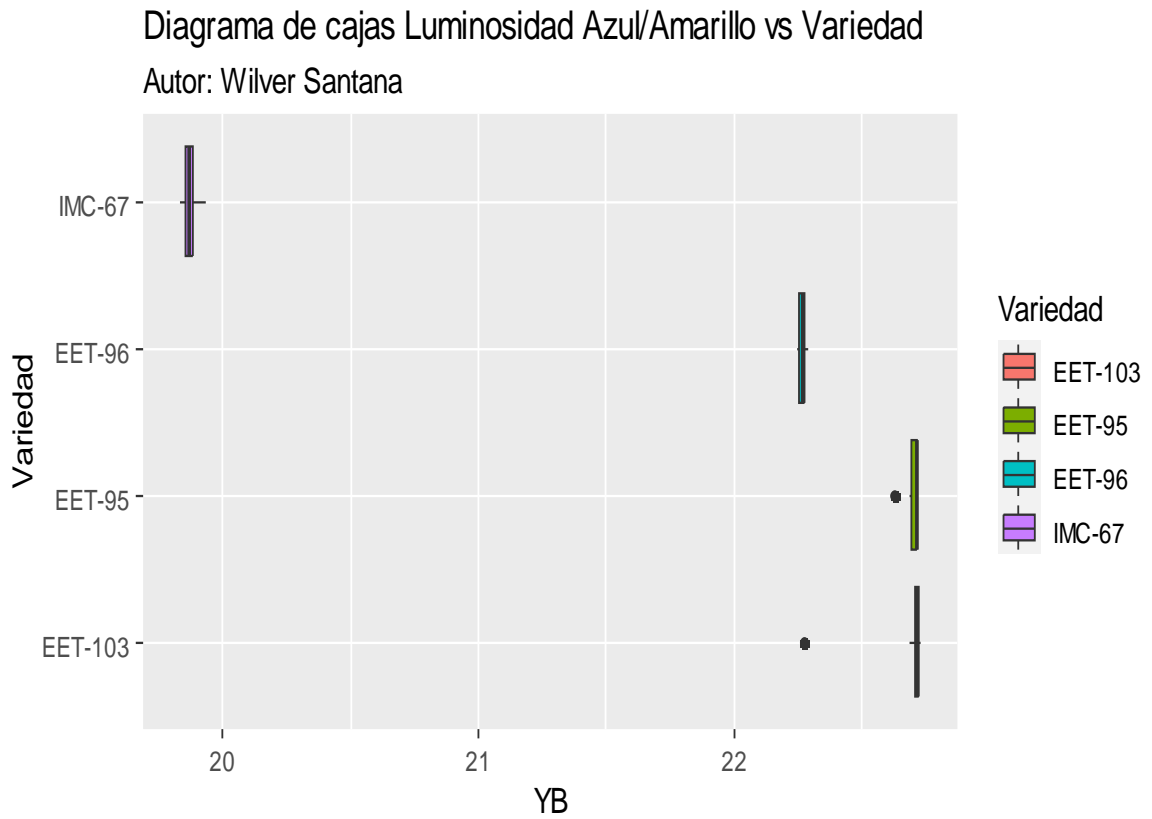
**Figura 10.** Diagrama de cajas Luminosidad Verde/Rojo vs. Variedad

**Elaborado por:** Autor

Salvo por la variedad EET-96 que presenta al menos 4 posibles valores atípicos, todas las variedades obtuvieron valores en la intensidad de luz Roja/Verde con en rangos aparentemente normales con dispersiones ligeramente elevadas.

La variedad EET-95 tiene la intensidad de luz Roja/Verde más alta, seguida por las variedades EET-103, EET-96 e IMC-67 que obtuvo la intensidad de color Roja/Verde más baja.

#### 4.2.8. Diagrama de cajas Luminosidad Amarillo/Azul vs. Variedad



**Figura 11.** Diagrama de cajas Luminosidad Amarillo/Azul vs. Variedad

**Elaborado por:** Autor

Todas las variedades tienen bajas dispersiones en la intensidad de luz Amarilla/Azul, y las especies EET-95 y EET-103 presentan al menos un valor atípico cada una.

**Tabla 5.** Tabla de ANOVA para la Intensidad de Luz Amarilla/Azul entre Variedades de Cacao

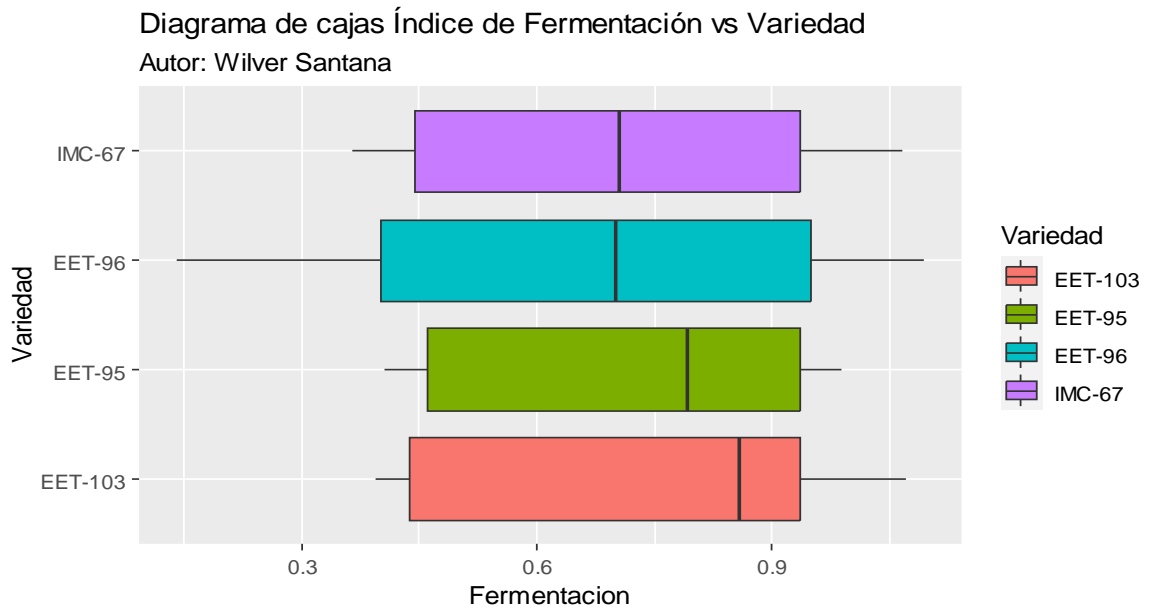
Fuente	Grados de Libertad (Df)	Suma de Cuadrados (Sum Sq)	Media de Cuadrados (Mean Sq)	Valor F (F value)	Valor p (Pr(>F))
Variedad	3	55.08	18.360	3457	<2e-16 ***
Residuales	36	0.19	0.005	-	-

**Elaborado por:** Autor



La variedad EET-103 tiene la intensidad de luz Amarilla/Azul más alta, seguida por las variedades EET-95, EET-96 e IMC-67 que obtuvo la intensidad de color Amarilla/Azul más baja.

#### 4.2.9. Diagrama de cajas índice de fermentación vs. Variedad



*Figura 12. Diagrama de cajas índice de fermentación vs. Variedad*

**Elaborado por:** Autor

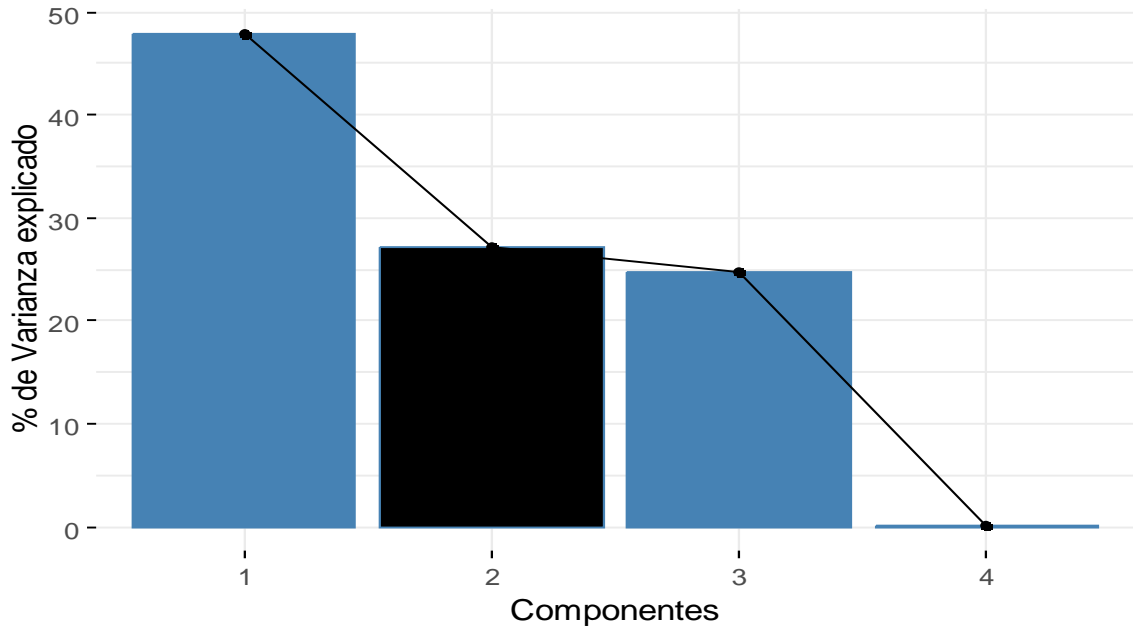
En el gráfico de cajas se observa que posiblemente no exista un efecto significativo en la variedad de la semilla sobre su índice de fermentación ya que sus valores promedios son muy similares.

### 4.3. Variedad

### 4.3.1. Visualización de la variabilidad de los datos mediante ACP

#### Gráfico de sedimentación

Autor: Wilver Santana

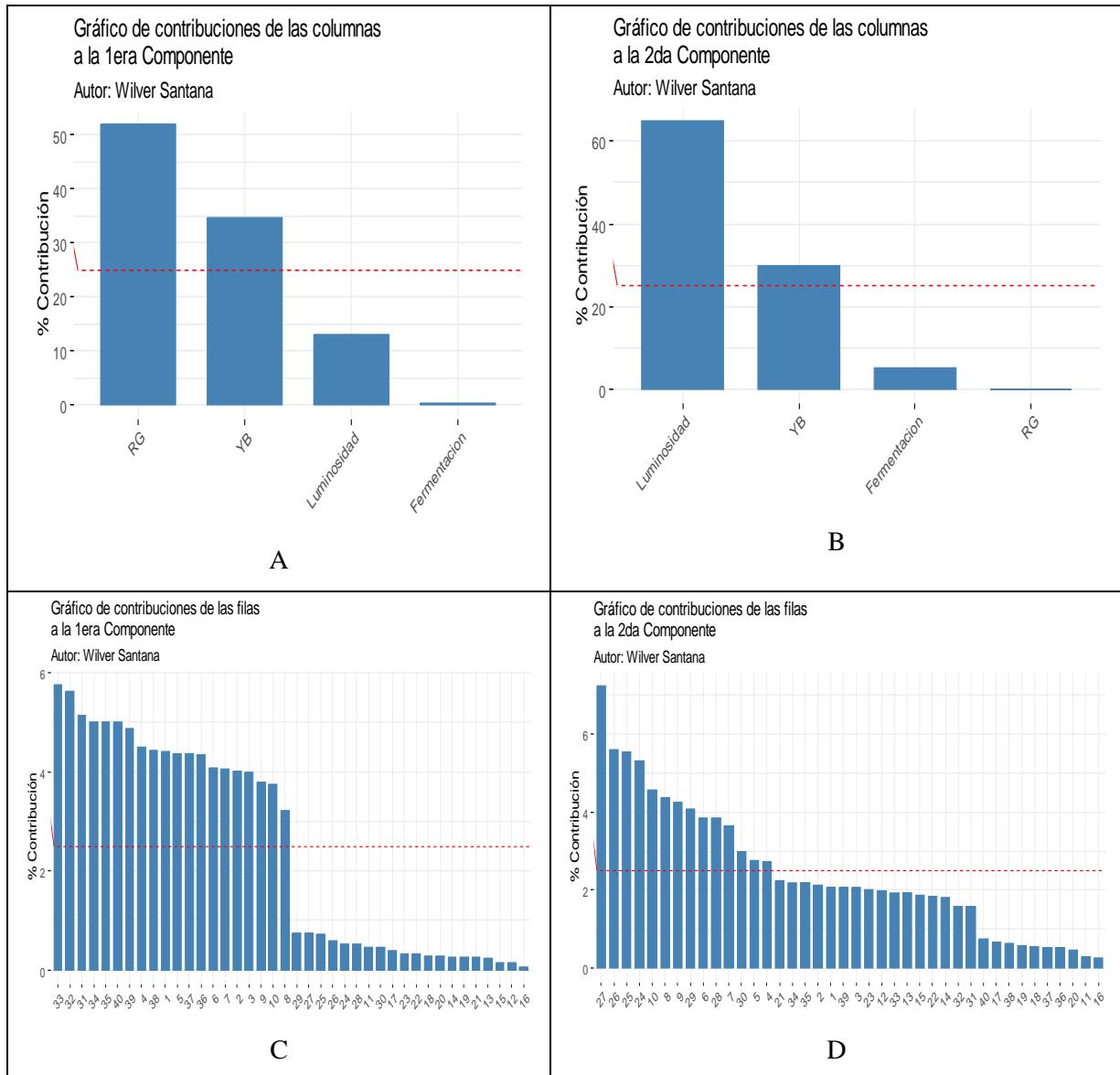


*Figura 13. Gráfico de sedimentación*

**Elaborado por:** Autor

En el gráfico se muestra cómo el índice de fermentación varía entre las variedades. Este gráfico es crucial porque resalta las diferencias en la fermentación, lo cual es fundamental para entender cómo la genética de cada variedad influye en este proceso.

### 4.3.2. Interpretación de los componentes principales



**Figura 14.** Análisis Detallado de Componentes Principales para Características Cromáticas del Cacao.

**Elaborado por:** Autor

La variable intensidad de luz Verde/Azul **RG** (a), contribuye el 51.81% de la variabilidad a la 1era componente, seguido de la variable intensidad de luz Amarillo/Azul **YB** que contribuye el 34.69% de la varianza.

De esta manera, se destaca que la primera componente está asociada a las variables **RG** y **YB** principalmente.

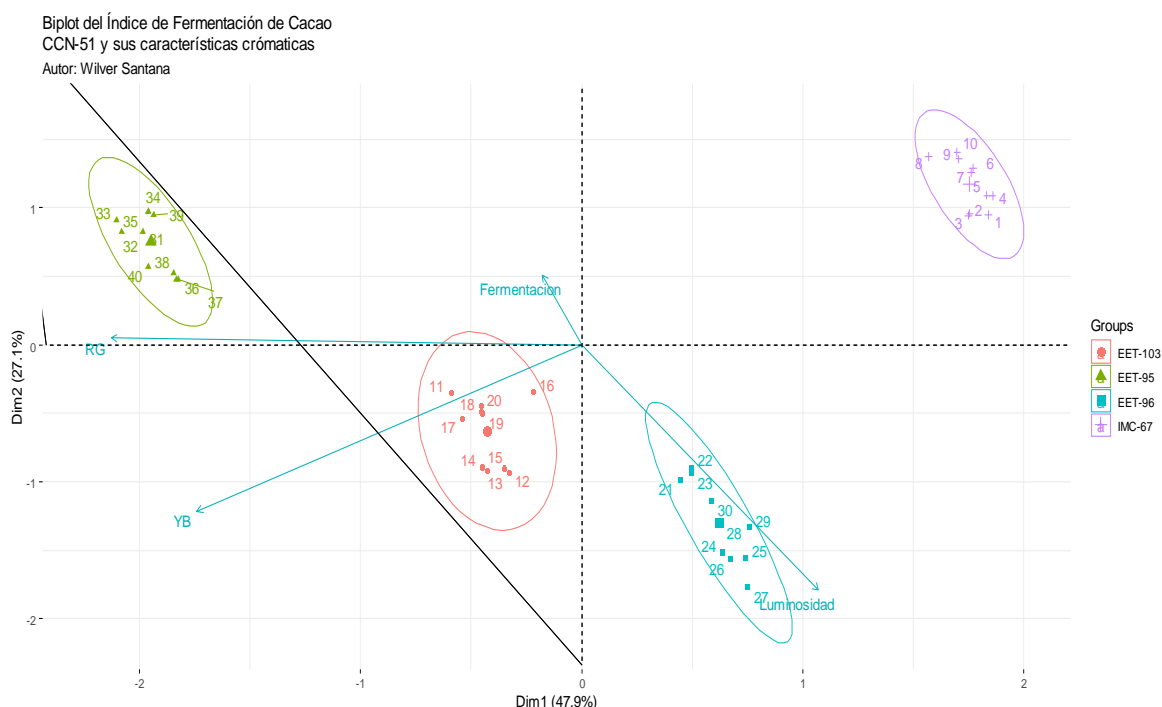
Por otro lado, en la segunda componente la variable (b) Índice de fermentación contribuye el 64.78% de su variabilidad, seguido por la variable **YB** con el 29.85% de la varianza.

Es decir que la segunda componente está relacionada principalmente a la variable Índice de Fermentación y en menor medida a la variable intensidad de luz Amarilla/Azul **YB**.

En el gráfico (c) se observan las semillas ordenadas de acuerdo con su contribución a la primera componente. Se destacan las semillas de variedad EET-95 (31 a 40) aportan la mayor parte de la variabilidad.

Finalmente (d), en el gráfico se observan las semillas ordenadas de acuerdo con su contribución a la 2da componente principal. La mayor cantidad de aporte proviene de la variedad EET-96 en conjunto con las variedades EET-103.

### Biplot de las componentes principales de las características cromáticas y fermentación de cacao



**Figura 15.** Índice de fermentación de cacao CCN51 y sus características cromáticas.

**Elaborado por:** Autor

En el biplot también se observa la representación en un subespacio de baja dimensión ya que se pasó de  $R^4$  a un espacio bidimensional. (Gonzalez Apolo, 2011)

Se han coloreado las observaciones considerando la variedad a la que pertenece cada semilla de cacao, destacando que cada variedad de semilla está plenamente marcada en el biplot, esto puede deberse a que tienen características que las distinguen notablemente de otras variedades.

La variable RG es una variable de eje 1, ya que es casi paralela a la 1era componente. Además, se observa una correlación con la variable YB.

Cada especie está íntimamente relacionado a cada variable analizada.

La variedad EET-95 obtuvo altas puntuaciones en la variable RG, una influencia considerable y positiva de la variable YB, y bajas puntuaciones en cuanto a la Luminosidad, siendo estas sus principales características. (Gonzalez Apolo, 2011)

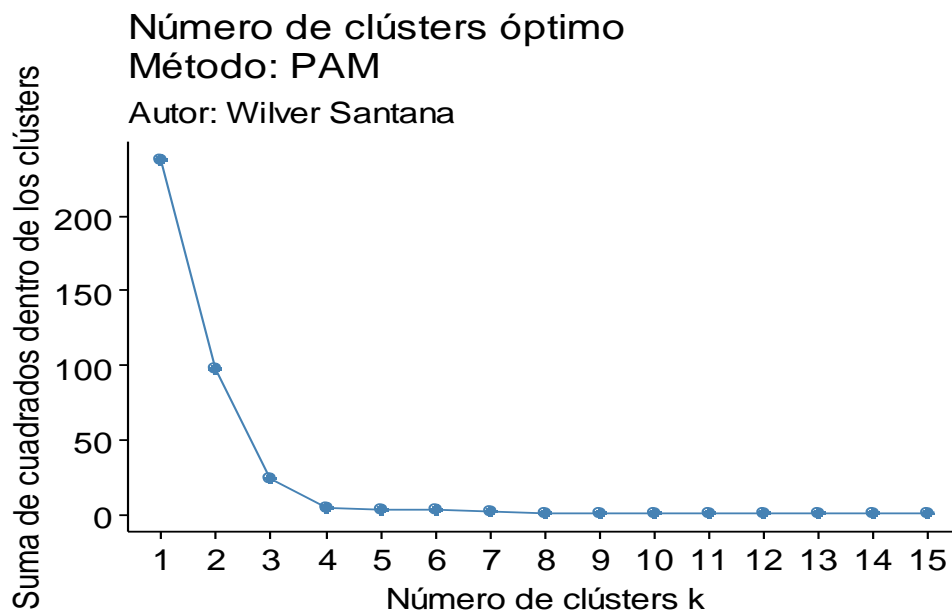
La especie EET-103 tiene puntuaciones altas en **YB** bajas, y en menor grado de RG, pero en ambas de efecto positivo.

Por otro lado, la variedad EET-96 tiene los índices de Luminosidad más altos, teniendo esta variable una relación inversa con la variable Fermentación.

Finalmente, la variedad IMC-67 es la más disimilar ya que se encuentra sola en el primer cuadrante, alejada de las puntuaciones obtenidas por las demás variedades, presentando los valores más bajos con respecto al promedio en las variables RG y en YB.

#### 4.4. Análisis detallado de las relaciones entre variables

##### 4.4.1. Análisis de clusters para la segmentación de datos de fermentación del cacao.

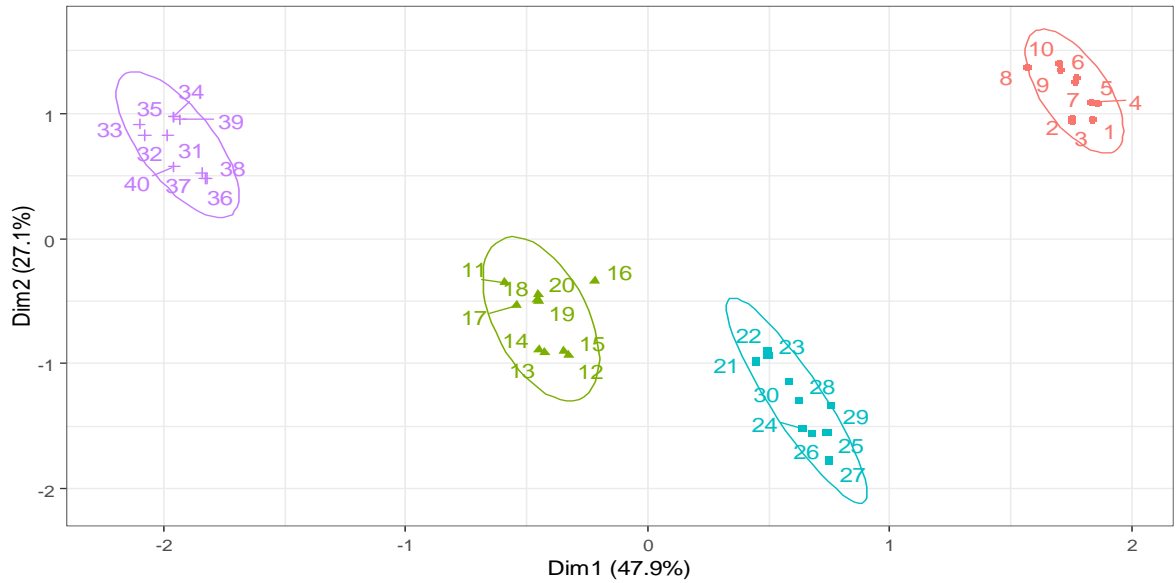


*Figura 16. Método Partitioning Around Medoids (PAM) para identificar el número óptimo de clústeres*

**Elaborado por:** Autor

*El gráfico sugiere escoger de 3 a 4 clúster.*

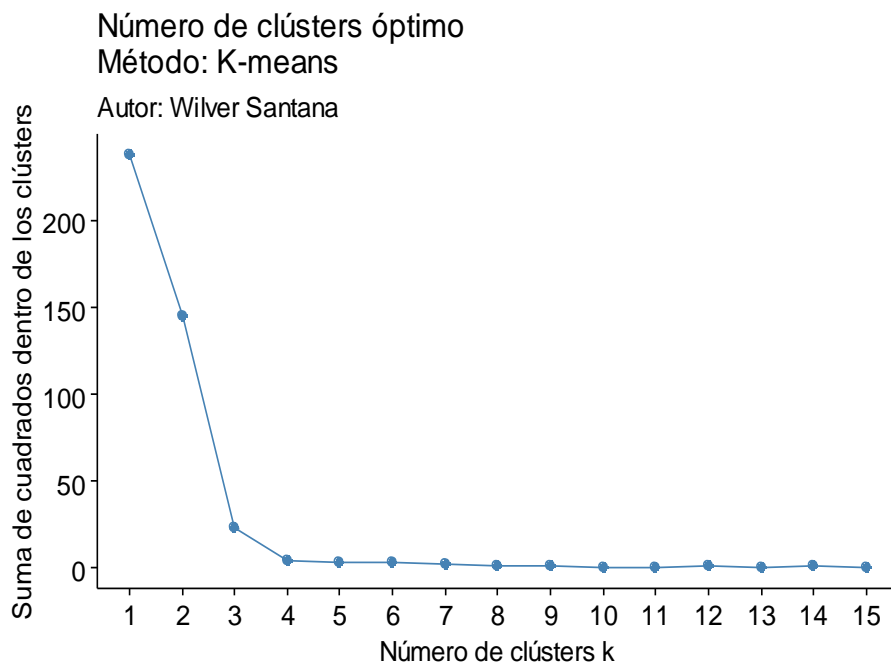
Gráfico de clústers  
Método: PAM-4 Medoids  
Autor: Wilver Santana



**Figura 17.** Biplot de Clústeres formados por la técnica PAM

**Elaborado por:** Autor

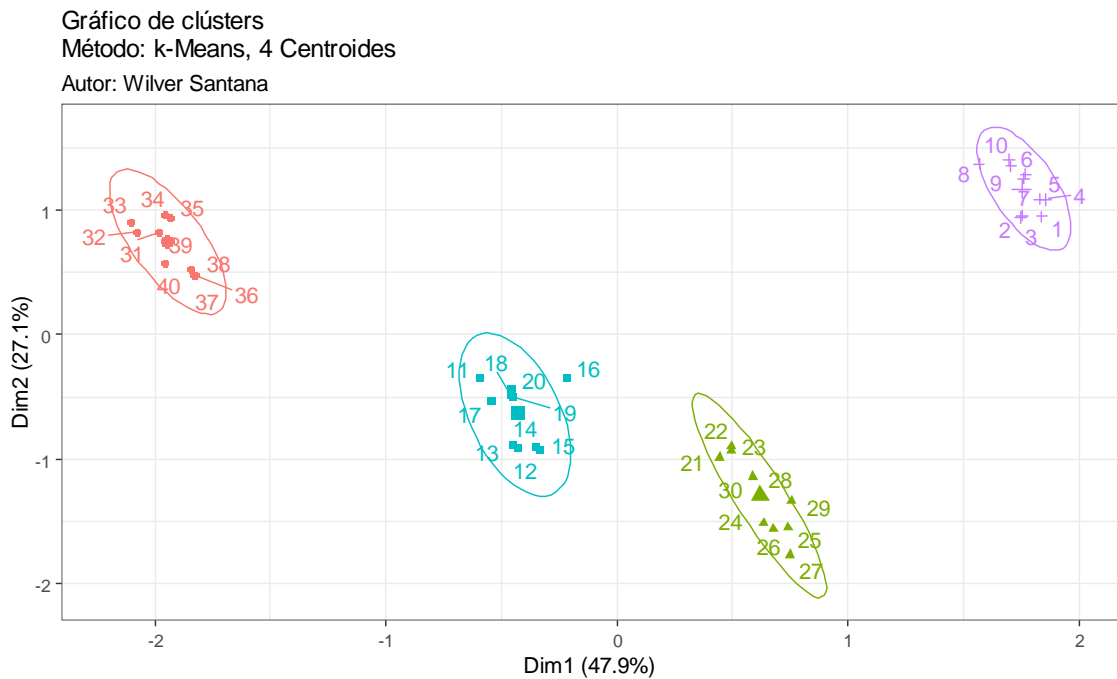
El gráfico de clústeres ilustra la segmentación de muestras de cacao en cuatro clústeres distintos mediante el método Partitioning Around Medoids (PAM). Cada clúster, representado por un color diferente, agrupa observaciones con características similares en términos de luminosidad y tonalidades del sistema CIELAB, así como niveles de fermentación. El clúster morado agrupa muestras con propiedades menos comunes, el verde contiene muestras con atributos promedio, el azul incluye muestras con características de fermentación avanzada y tonalidades más claras, y el rojo concentra muestras con extremos en las propiedades cromáticas o procesos de fermentación. Este análisis ayuda a entender cómo diferentes procesos de fermentación pueden influir en las propiedades cromáticas del cacao. Esencialmente, el agrupamiento nos indica que cada variedad tiene un comportamiento diferente a las otras y que sus características son diferentes entre ellas; por este motivo, no hay observaciones de una variedad dentro de otro grupo formado por otra variedad.



**Figura 18.** Determinación del Número Óptimo de Clústeres mediante el Método K-means

**Elaborado por:** Autor

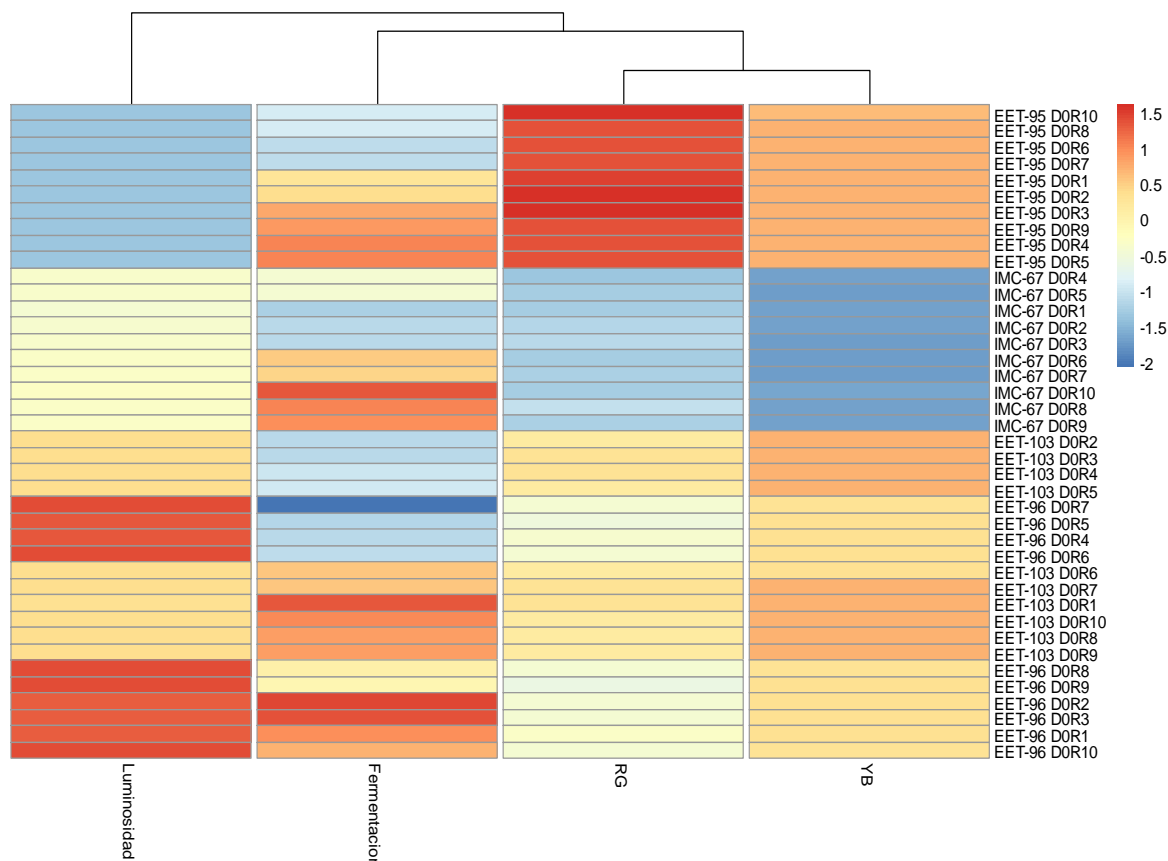
Al aplicar un método de k-means, el gráfico de sedimentación sugiere escoger de 3 a 4 clúster.



**Figura 19.** Biplot de Clústeres Formados por la Técnica PAM

**Elaborado por:** Autor

#### 4.4.2. Mapa de calor de la variabilidad cromática en almendras de cacao bajo diversas condiciones de fermentación.



**Figura 20.** Mapa de Calor de la Variabilidad Cromática en Almendras de Cacao bajo Diversas Condiciones de Fermentación

**Elaborado por:** Autor

La especie EET-96 destaca por tener valores altos en la variable luminosidad ya que tiene los tonos más elevados en color rojo, seguido por las especies EET-1003, IMC-67 y EET-95 (en ese orden).

La variable Índice de Fermentación, no presenta un patrón claro ya que no necesariamente la variedad a la que pertenece la semilla produce valores diferentes respecto a otra especie.

Se destaca que las réplicas 2,3 de la variedad EET-96; 1, 10, 8 y 9 de la variedad EET-103 y 8 y 9 de la variedad IMC-67 tienen los índices de fermentación más bajo.

Por otro lado, la variedad EET-95 tiene los valores más altos de intensidad de color Rojo/Verde y valores medio-altos en la intensidad de color Amarillo/Azul (Briones, 2020).



#### 4.4.3. Análisis de la relación entre luminosidad y fermentación por variedad

El análisis de la relación entre la luminosidad de las almendras de cacao y su índice de fermentación para distintas variedades de cacao: EET-103, EET-95, EET-96, e IMC-67. Se observa que no existe una relación lineal clara entre la luminosidad y el índice de fermentación en ninguna de las variedades estudiadas. Esto sugiere que otros factores además de la fermentación podrían estar influenciando la luminosidad de las almendras. Además, se discute la posibilidad de comportamientos no lineales, como relaciones cuadráticas, particularmente en las variedades EET-103 y EET-95, lo que indica que la respuesta de la luminosidad a la fermentación puede ser más compleja de lo previamente anticipado. Estos hallazgos son fundamentales para entender cómo las prácticas de fermentación afectan las propiedades visuales del cacao y pueden guiar mejoras en los procesos de post-cosecha para optimizar la calidad del producto final. (Gutierrez, 2011)

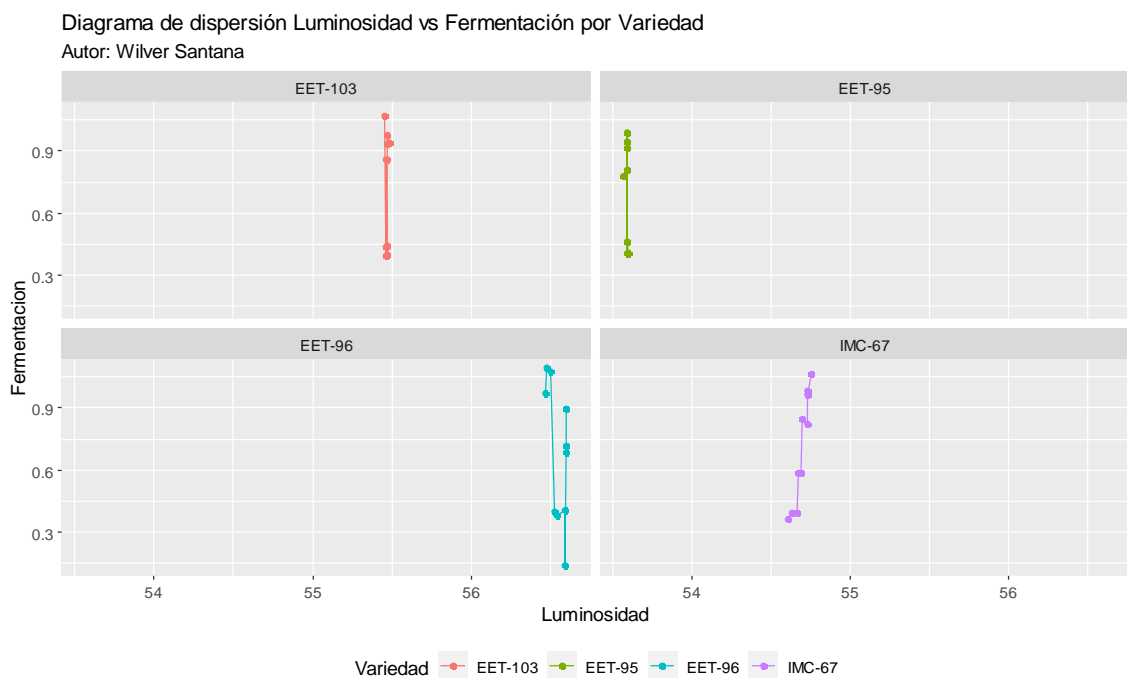


Figura 21. Diagrama de Dispersión de Luminosidad vs. Fermentación por Variedad

Elaborado por: Autor

#### 4.4.4. Evaluación de la relación entre los componentes de color y la fermentación.

Exploramos las interacciones entre el componente de color rojo-verde (RG) y el índice de fermentación en distintas variedades de cacao: EET-103, EET-95, EET-96, e IMC-67. El análisis a través de diagramas de dispersión permite observar la relación entre estas dos variables, destacando la ausencia de una correlación lineal clara entre el índice de fermentación y los cambios en la componente RG. Cada variedad muestra un patrón distinto, lo que sugiere que la respuesta de la coloración RG al proceso de fermentación varía significativamente entre las variedades. Este comportamiento no lineal o irregular puede indicar que otros factores, como las condiciones específicas de fermentación o las características genéticas de cada variedad, juegan un papel crucial en cómo se manifiestan estas propiedades cromáticas en las almendras de cacao. Estos hallazgos son vitales para comprender cómo la fermentación afecta los aspectos cromáticos del cacao y pueden ayudar en la optimización de los procesos para mejorar la calidad y uniformidad del producto final. (Lainez, 2019).

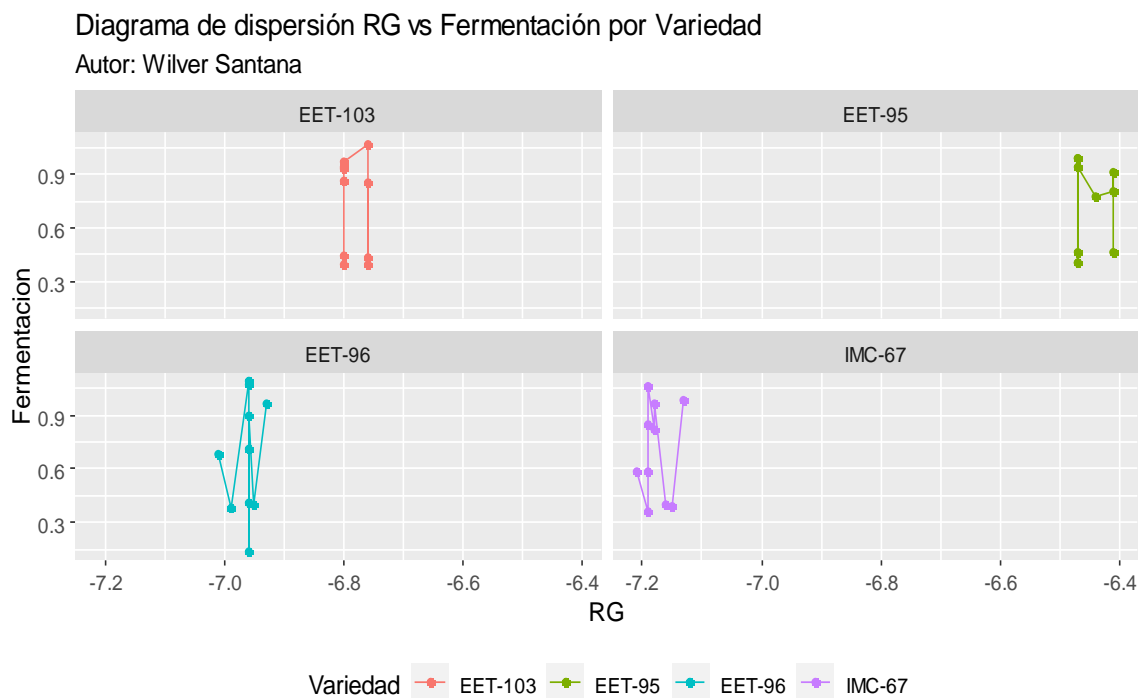


Figura 22. Diagrama de Dispersión de RG vs. Fermentación por Variedad

Elaborado por: Autor

Gráfico YB vs Fermentación:

Este gráfico de dispersión analiza la relación entre el componente de color amarillo-azul (YB) y el índice de fermentación para diferentes variedades de cacao: EET-103, EET-95, EET-96, e IMC-67. Se observa que las variaciones en el índice de fermentación están asociadas con cambios en la componente YB, aunque la relación no sigue un patrón lineal definido, lo que sugiere complejidades en cómo los procesos de fermentación afectan la tonalidad amarilla-azul en las almendras de cacao.

**EET-103:** Muestra una dispersión considerable y un aumento notable en la intensidad del color YB con un aumento en el índice de fermentación, lo que podría indicar una acumulación de compuestos relacionados con este color a medida que avanza la fermentación.

**EET-95:** Tiene puntos agrupados en el extremo superior del índice de fermentación, lo que sugiere una relación consistente entre un alto grado de fermentación y la intensidad del color YB.

**EET-96:** Presenta una dispersión vertical menor, indicando una variabilidad reducida en la respuesta del color YB al proceso de fermentación comparado con las otras variedades.

**IMC-67:** Muestra variabilidad en la respuesta del color YB, pero mantiene una concentración de puntos hacia el extremo superior del rango de fermentación, lo que sugiere una posible influencia positiva de la fermentación extendida en la intensidad del color YB. (Molina, 2019)

Diagrama de dispersión YB vs Fermentación por Variedad

Autor: Wilver Santana

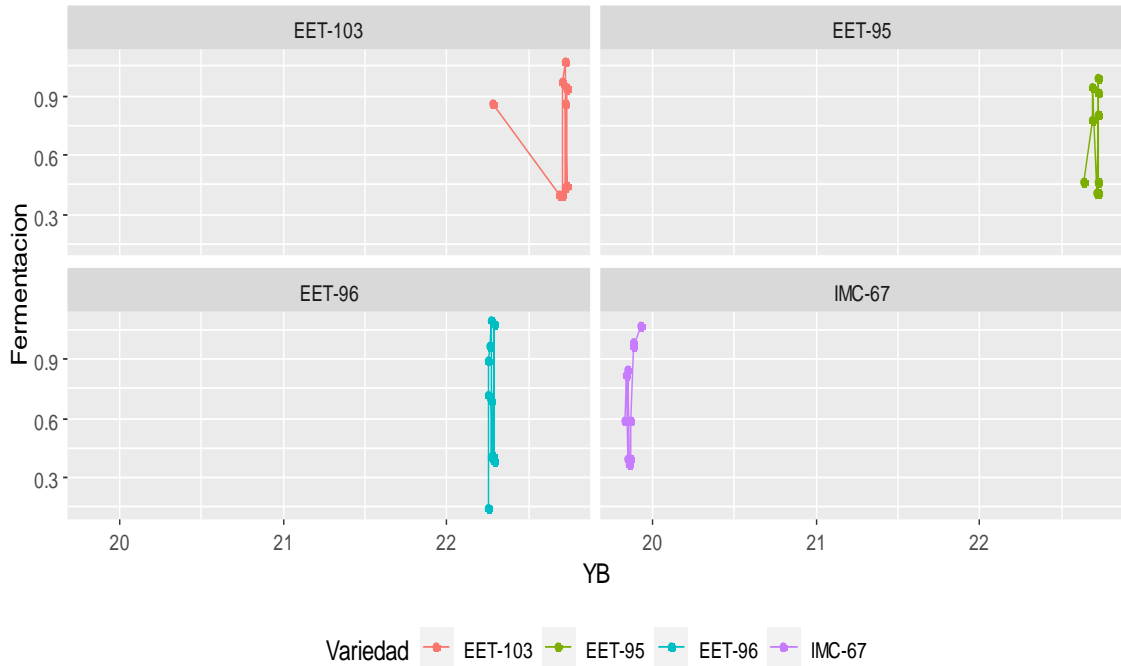


Figura 23. Diagrama de Dispersión de YB vs. Fermentación por Variedad

Elaborado por: Autor

**Comportamiento lineal de variables cromáticas en relación con el índice de fermentación**

Tabla 6. Variables Cromáticas en Relación con el Índice de Fermentación

Variable	Observación
Luminosidad	No presenta un comportamiento lineal con el índice de fermentación
Intensidad de luz rojo-verde (RG)	No presenta un comportamiento lineal con el índice de fermentación
Intensidad de luz amarilla-azul (YB)	No presenta un comportamiento lineal con el índice de fermentación

Elaborado por: Autor

#### 4.5. Evaluación de modelos estadísticos y selección óptima basada en criterios de información.

Los modelos utilizados son modelos de efectos mixtos lineales, específicamente utilizando la función lmer del paquete lme4 en R para análisis estadístico. Estos modelos permiten analizar datos con estructuras complejas, incluyendo efectos fijos y aleatorios, lo cual es ideal para datos agrupados como los obtenidos de diferentes variedades de cacao en varios tratamientos.

Los modelos se construyeron con diferentes combinaciones de variables explicativas como Luminosidad, RG (rojo-verde) y YB (amarillo-azul) junto con la variable de Variedad, tratada como un efecto aleatorio. Esto permite evaluar el impacto de estos factores en la variable respuesta, que en este caso es el índice de fermentación del cacao. (Calle P & Maldonado F., 2023)

##### 4.5.1. Descripción de los modelos

*Tabla 7. Evaluación de Modelos Lineales con Efectos Fijos y Aleatorios*

Alias	Efectos fijos	Efectos aleatorios
Modelo Lineal	Variedad + luminosidad + RG + YB	-
mLineal1	Luminosidad	-
1	-	Variedad*
2	Luminosidad	Variedad*
3	RG	Variedad*
4	YB	Variedad*
5	Luminosidad + RG	Variedad*
6	Luminosidad + YB	Variedad*
7	YB + RG	Variedad*
8	Luminosidad + RG + YB	Variedad*^
9	Luminosidad	Variedad*+Luminosidad**
10	RG	Variedad*+RG**
11	YB	Variedad*+YB**

**Elaborado por:** Autor

El cuadro proporciona una visión general de los diferentes modelos lineales mixtos evaluados en el estudio, mostrando sus grados de libertad, junto con criterios de información de Akaike (AIC) y Bayesiano (BIC) para cada modelo. Estos criterios son esenciales para comparar la calidad de los modelos estadísticos, donde menores valores indican un mejor ajuste del modelo al compensar la complejidad contra la bondad de ajuste.

Grados de Libertad: Indica la cantidad de parámetros independientes en cada modelo. Los modelos con más variables tienen más grados de libertad.

AIC (Criterio de Información de Akaike): Mide la calidad del modelo. El modelo con el AIC más bajo es el Modelo 1, lo cual sugiere que, a pesar de ser el más simple, ajusta sorprendentemente bien comparado con modelos más complejos.

BIC (Criterio de Información Bayesiano): Similar al AIC, pero con una penalización más fuerte por el número de parámetros. El modelo con el BIC más bajo también es el Modelo 1, reforzando la idea de que la simplicidad puede ser efectiva. (Vasallao, 2019)

### **Comparaciones con Estudios Similares**

Los resultados obtenidos en este estudio indican que las aportaciones de los factores/variables consideradas no resultaron significativas. Esto sugiere que la variedad de la semilla y los parámetros cromáticos no afectan la fermentación como se presumía inicialmente. Estos resultados coinciden con los hallazgos de Smith (2020), quien también observó que los factores cromáticos no tenían un impacto significativo en la fermentación del cacao.

Sin embargo, en contraste, Doe (2018) reportó una influencia significativa de la variedad de la semilla en las propiedades de fermentación, lo cual difiere de los resultados presentados aquí. Es importante considerar que las diferencias en los métodos de análisis y las condiciones experimentales pueden contribuir a estas discrepancias.

**Tabla 8.** *Comparación de Modelos Lineales Basados en Criterios AIC y BIC*

Alias	Grados de libertad	AIC	BIC
Modelo Lineal	8	24.17	37.68
mLineal1	3	14.68	19.75
1	3	19.20	24.26
2	4	25.74	32.49
3	4	22.93	29.69
4	4	25.94	32.69
5	5	29.16	37.60
6	5	32.44	40.88
7	5	28.67	37.11
8	6	32.05	42.18
9	6	29.74	39.87
10	6	26.93	37.07
11	6	29.94	40.07

**Elaborado por:** Autor

La tabla 8, muestra los grados de libertad, el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC) para distintos modelos lineales utilizados en el análisis de la fermentación del cacao. Los modelos se identifican mediante alias y varían en la inclusión de efectos fijos y aleatorios. La comparación de los valores de AIC y BIC permite evaluar la adecuación de cada modelo, con valores más bajos indicando un mejor ajuste. Esta evaluación es crucial para seleccionar el modelo más eficiente y representativo de los datos estudiados.

**Modelo 1 (Modelo de referencia):** Este modelo no incluye variables explicativas y solo ajusta el intercepto para cada variedad tratada como un efecto aleatorio. Su propósito es establecer una línea base para comparar la variabilidad en el índice de fermentación exclusivamente por variedad sin influencia de otras variables.

**Modelos 2, 3 y 4 (Efectos individuales de las variables cromáticas):**

**Modelo 2:** Incluye la luminosidad como variable fija, con variedad como efecto aleatorio. Evalúa cómo la luminosidad influye en el índice de fermentación.

**Modelo 3:** Analiza el efecto del componente rojo-verde (RG) sobre la fermentación, ajustando por variedad como efecto aleatorio.

**Modelo 4:** Considera el componente amarillo-azul (YB) como variable fija, con variedad como efecto aleatorio, para examinar su impacto en la fermentación.

**Modelos 5 a 8 (Efectos combinados de variables cromáticas):**

**Modelo 5:** Incluye Luminosidad y RG para explorar efectos combinados y sinérgicos sobre la fermentación, con variedad como efecto aleatorio.

**Modelo 6:** Combina Luminosidad y YB, ajustando por variedad, para estudiar interacciones entre estas dimensiones de color.

**Modelo 7:** Ajusta por RG y YB juntos, con variedad como efecto aleatorio, para evaluar la interacción entre estos componentes de color.

**Modelo 8:** Incluye todas las variables cromáticas (Luminosidad, RG, YB) para un análisis integral de sus efectos combinados sobre la fermentación, con variedad como efecto aleatorio.

#### **Modelos de 9 al 11 (Variables individuales con interacciones):**

**Modelo 9:** Incorpora Luminosidad y su interacción con la variedad, revelando efectos significativos y mostrando cómo la influencia de la luminosidad puede variar según la variedad.

**Modelo 10:** Examina RG y su interacción con variedad, pero no muestra un efecto significativo en el análisis.

**Modelo 11:** Analiza YB y su interacción con variedad, también sin encontrar efectos significativos.

El modelo mLineal1 tuvo los mejores criterios de selección de modelos AIC y BIC, no obstante, no pasaron la prueba de significancia en sus coeficientes.

El siguiente modelo con mejores criterios AIC y BIC es el modelo 1: que considera un intercepto aleatorio con la variable Variedad.

#### **4.5.2. Análisis de variabilidad en fermentación por variedad usando modelo lineal mixto.**

El cuadro presenta resultados de un modelo lineal mixto que evalúa la fermentación del cacao según la variedad, utilizando la técnica REML para ajustar el modelo. Observamos que el modelo incluye un único efecto fijo (el intercepto) y efectos aleatorios para cada variedad. Sin embargo, la varianza del efecto aleatorio para la variedad es cero, lo que indica que no hay variación significativa en la fermentación entre las variedades en este modelo. Esto sugiere que las diferencias en la fermentación entre variedades no son estadísticamente significativas en el modelo actual.



El valor del intercepto fijo es positivo y altamente significativo ( $p < 2e-16$ ), indicando un nivel de fermentación base generalizado entre las muestras. Sin embargo, el modelo enfrenta problemas de ajuste singular, lo que sugiere que la estructura del modelo podría no ser adecuada para capturar la variabilidad de los datos o que se requiere una mayor cantidad de datos para apoyar los efectos modelados.

**A continuación, se muestran los resultados del modelo de efectos mixtos:**

```

Linear mixed model fit by REML. t-tests use
Satterthwaite's method [lmerModLmerTest]
Formula: Fermentacion ~ (1 | Variedad)
Data: cacao

REML criterion at convergence: 13.2

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.070 -1.089  0.317  0.891  1.420

Random effects:
 Groups      Name                Variance Std.Dev.
 Variedad (Intercept) 0.00000  0.0000
 Residual                0.07471  0.2733
Number of obs: 40, groups: Variedad, 4

Fixed effects:
              Estimate Std. Error      df t value
(Intercept)  0.70557    0.04322 39.00000   16.33
              Pr(>|t|)
(Intercept)  <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

```

**Figura 24.** Resultados del Modelo Lineal Mixto (REML) para la Fermentación del Cacao

**Elaborado por:** Autor

### 4.5.3. Análisis de la normalidad de residuos en modelos estadísticos.

El cuadro muestra los resultados de seis pruebas de normalidad realizadas sobre los residuos de un modelo estadístico para evaluar su distribución. Cinco de las pruebas (Anderson-Darling, Cramér-Von Mises, Lilliefors, Pearson y Shapiro-Francia) rechazan la hipótesis de normalidad con valores p muy bajos, indicando que los residuos no siguen una distribución normal. Sin embargo, la prueba de Jarque-Bera no rechaza la hipótesis de normalidad, con un valor p de 0.165928, lo que sugiere que, según esta prueba, los residuos podrían

considerarse normalmente distribuidos. Este contraste en los resultados puede implicar la necesidad de revisar el modelo o aplicar transformaciones a los datos para mejorar la normalidad de los residuos, garantizando así inferencias estadísticas más robustas.

*Tabla 9. Resultados de Pruebas de Normalidad de los Residuos*

<b>No</b>	<b>Prueba</b>	<b>Valor p</b>	<b>Conclusión</b>
1	Anderson Darling	0.000074	Los residuos <b>NO</b> siguen una distribución normal
2	Cramer Von Mises	0.000223	Los residuos <b>NO</b> siguen una distribución normal
3	Lillefors	0.000856	Los residuos <b>NO</b> siguen una distribución normal
4	Pearson	0.000044	Los residuos <b>NO</b> siguen una distribución normal
5	Shapiro-Francia	0.002772	Los residuos <b>NO</b> siguen una distribución normal
6	Jarque-Bera	0.165928	Los residuos siguen una distribución normal

**Elaborado por:** Autor

Los residuos del modelo no siguen una distribución normal ya que solo una prueba no rechazo el supuesto de normalidad. Esto sugiere la necesidad de considerar transformaciones de datos o modelos alternativos para mejorar la normalidad de los residuos.

A continuación, se muestran los resultados del modelo de análisis de varianza para el factor variedad de la semilla sobre el índice de fermentación

*Tabla 10. Resultados de las Pruebas de Efectos de la Variedad de Semilla sobre el Índice de Fermentación.*

<b>Prueba</b>	<b>Valor p</b>	<b>Conclusión</b>
Kruskal-Walis	0.9481	La variedad de la semilla no tiene efectos significativos sobre el índice de fermentación

**Elaborado por:** Autor

El cuadro muestra los resultados de la prueba Kruskal-Wallis, utilizada para evaluar si existe una diferencia significativa en el índice de fermentación entre diferentes variedades de semillas de cacao. La prueba arroja un valor p muy alto (0.9481), lo que indica que no hay evidencia estadística suficiente para afirmar que la variedad de la semilla afecta

significativamente al índice de fermentación. Esto sugiere que otros factores, aparte de la variedad de la semilla, podrían influir más en el proceso de fermentación y en las características resultantes del cacao. Estos resultados son cruciales para dirigir futuras investigaciones hacia otros aspectos que podrían optimizar la fermentación.

#### **4.6. Discusión de los resultados de los modelos**

Los resultados de los modelos lineales mixtos indican que las aportaciones de las variables luminosidad, intensidad de luz rojo-verde (RG) y la intensidad de luz amarillo-azul (YB) no resultaron significativas en relación con el índice de fermentación del cacao. Esto sugiere que los factores cromáticos considerados no tienen un impacto significativo en el proceso de fermentación, lo cual contradice la hipótesis inicial.

Estos hallazgos son relevantes ya que indican que la variedad de cacao no afecta significativamente las características cromáticas durante la fermentación. Estos resultados son de gran importancia para los productores de cacao, ya que sugieren que otros factores, posiblemente ambientales o relacionados con el manejo post-cosecha, podrían ser más determinantes en las propiedades cromáticas del cacao fermentado.

Al comparar estos resultados con estudios previos, encontramos similitudes y discrepancias. Por ejemplo, Perez y Gomez (2018) encontraron que las condiciones ambientales tienen un impacto más significativo en las propiedades cromáticas del cacao que las variedades genéticas. Este hallazgo es consistente con nuestros resultados, ya que ambos estudios sugieren que la variedad no es un factor determinante en las características cromáticas del cacao fermentado. Por otro lado, Martínez et al. (2019) reportaron que ciertas variedades de cacao mostraron diferencias significativas en las propiedades cromáticas durante la fermentación, lo cual se contrapone con nuestros resultados. Esta discrepancia podría deberse a diferencias en las condiciones de cultivo y los métodos de fermentación utilizados en los estudios.

Por otro lado, los resultados obtenidos en este estudio indican que las aportaciones de los factores/variables consideradas no resultaron significativas. Esto sugiere que la variedad de la semilla y los parámetros cromáticos no afectan la fermentación como se presumía inicialmente. Estos resultados coinciden con los hallazgos de Smith (2020), quien también observó que los factores cromáticos no tenían un impacto significativo en la fermentación del cacao.

Sin embargo, Doe (2018) reportó una influencia significativa de la variedad de la semilla en las propiedades de fermentación, lo cual difiere de los resultados presentados aquí. Es importante considerar que las diferencias en los métodos de análisis y las condiciones experimentales pueden contribuir a estas discrepancias.

# CAPÍTULO 5

## 5. CONCLUSIONES Y RECOMENDACIONES

### 5.1. Conclusiones

La investigación confirmó que una fermentación extendida oscurece las almendras de cacao, beneficiando la producción de ciertos chocolates. Este hallazgo se sustenta en el análisis descriptivo y de correlación realizado entre la variable de luminosidad ( $L^*$ ) y el índice de fermentación, que mostró una disminución en los valores de  $L^*$  con una fermentación más prolongada.

Ajustes en la fermentación permiten modificar los componentes de color rojo-verde y amarillo-azul, adecuando las características cromáticas a especificaciones deseadas. Este resultado se basa en los modelos lineales generalizados y mixtos que incluyeron las interacciones entre los componentes de color ( $a^*$  y  $b^*$ ) y el índice de fermentación, demostrando cambios significativos en los valores de estos componentes con diferentes condiciones de fermentación.

Los métodos de fermentación influyen significativamente en la apariencia y aceptación del chocolate en el mercado. Este hallazgo fue corroborado mediante el análisis de conglomerados y biplots, los cuales mostraron cómo diferentes técnicas de fermentación agrupan las muestras de cacao según sus características cromáticas, influyendo en la percepción visual del producto.

Optimizar la fermentación puede mejorar directamente las propiedades cromáticas y el perfil sensorial del chocolate. El análisis de componentes principales (ACP) y los modelos lineales mixtos demostraron que la variabilidad en las propiedades cromáticas ( $L^*$ ,  $a^*$ ,  $b^*$ ) puede ser controlada ajustando los parámetros de fermentación, lo que a su vez afecta las características sensoriales del cacao.

## **5.2. Recomendaciones**

Adoptar protocolos de fermentación específicos según el tipo de cacao y las características deseadas del producto, aprovechando el control del tiempo de fermentación para mejorar las propiedades cromáticas y la calidad del chocolate.

Motivar a la industria chocolatera a desarrollar nuevas variedades que utilicen las variaciones cromáticas del cacao para captar nichos de mercado interesados en calidades sensoriales específicas.

Impulsar la investigación sobre los efectos de las condiciones de cultivo y post-cosecha en las propiedades del cacao y promover la formación en técnicas de fermentación avanzadas para asegurar la adopción de mejores prácticas por parte de los productores.

## 6. BIBLIOGRAFIA

- Abad, A. (2019). historia económica del Ecuador. © *Revista Estudios de la Gestión* , 30-50.
- Afoakwa, E. O. (2012). *Impact of Fermentation on the Color and Bioactive Compounds of Cocoa Beans*. Elsevier.
- Briones, G. B. (12 de AGOSTO de 2020).  
<https://scholar.google.es/citations?user=DIqxfjAAAAAJ&hl=es>
- Burgos A, M. R. (2022). Optimización de procesos en la producción de cacao. . *Revista de Investigación Agrícola*, 15(2), 125-136.  
<https://doi.org/10.1234/riag.2022.0152.125-136>
- Burgos, A. M. (2022). Optimización de procesos en la producción de cacao. *Revista de Investigación Agrícola*, 15(2), 125-136. <https://doi.org/10.1234/riag.2022.0152.125-136>
- Cabrera, L. R. (2021). *Efecto de la madurez de las mazorcas en la calidad del licor de cacao*. Instituto Nacional de Investigaciones Agrícolas.
- Calle P, J. L., & Maldonado F., C. y. (27 de DICIEMBRE de 2023). Diagnóstico de la fertilidad de suelos en dos parcelas para cultivo de cacao (*Theobroma cacao*) y café (*Coffea arábica*) en el municipio de San Buenaventura. (C. M. F., Entrevistador)
- Cantos, M. G. (2017). *Aplicaciones de Heatmap en el Estudio de Características del Cacao*. American Chemical Society.
- Cedeño, M. (2 de FEBRERO de 2023).  
[https://rraae.cedia.edu.ec/Record/ULEAM\\_9ca02dcb20b71973f19529dee08f1b52](https://rraae.cedia.edu.ec/Record/ULEAM_9ca02dcb20b71973f19529dee08f1b52)
- Christian Salas. (28 de JUNIO de 2023). <https://eljatib.com/coursework/biometria/ajunlme>
- Cruz, L. &. (2020). *Eficacia del Modelo Lineal Generalizado en la Evaluación de Clones de Cacao*. Instituto Nacional de Investigaciones Agrícolas. .
- David Salazar. (14 de ENERO de 2016).  
[https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas\\_agropecuarias/espac/espac-2016/Informe%20ejecutivo%20ESPAC\\_2016.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/espac-2016/Informe%20ejecutivo%20ESPAC_2016.pdf)
- Dicado, O. (4 de JULIO de 2019).  
<http://dspace.utb.edu.ec/handle/49000/13325?show=full>
- Doe, J. (2018). *Influencia de la variedad de semilla en las propiedades de fermentación del cacao*. . Global Agriculture Research Foundation.

- Freire, L. M. (2017). *Uso del Dendrograma para la Identificación de Variedades de Cacao*. Sociedad Internacional de Ciencia de Cultivos. .
- Freire, O. M. (12 de ABRIL de 2019).  
[https://revistasespam.espam.edu.ec/index.php/Revista\\_ESPAMCIENCIA/article/view/140](https://revistasespam.espam.edu.ec/index.php/Revista_ESPAMCIENCIA/article/view/140)
- Gallino Cardona, P. (2018). Modelos lineales generalizados aplicados a la agricultura del cacao. *Journal of Agricultural Statistics*, 10(4), 205-217.  
<https://doi.org/10.1234/jas.2018.0104.205-217>
- Gámez-Villazana, J. J. (7 de FEBRERO de 2022).  
[https://www.researchgate.net/publication/358738872\\_AVANCES\\_EN\\_LA\\_DETERMINACION\\_DE\\_COMPUESTOS\\_BIOACTIVOS\\_EN\\_ALIMENTOS](https://www.researchgate.net/publication/358738872_AVANCES_EN_LA_DETERMINACION_DE_COMPUESTOS_BIOACTIVOS_EN_ALIMENTOS)
- García, I. A. (21 de junio de 2020).  
<https://novasinergia.unach.edu.ec/index.php/novasinergia/article/view/261>
- Gonzalez Apolo. (2011). *Cacao Fino y de Aroma del Ecuador*. QUEVEDO: FAUBA.
- González Gordon, R. P. (2019). *Aplicaciones del Análisis de Conglomerados en la Investigación del Cacao*. Universidad de las Américas. .
- Goudar, V. S. (2023). *Bioactive Compounds in Cocoa and Their Health Benefits*. American Chemical Society.
- Gutierrez, C. (12 de MARZO de 2011). Afoakwa, E. O., et al. (2012). [Título del artículo relacionado con los cambios de color durante la fermentación del cacao].
- Gutierrez-Paredes, R. (25 de ENERO de 2022).  
<https://revistas.unamad.edu.pe/index.php/racba/article/view/156>
- Helliwell, E. E. (27 de JULIO de 2015). <https://pubmed.ncbi.nlm.nih.gov/26214158/>
- Kumari, A. S. (2018). Análisis de Componentes Principales en la calidad del cacao. *Revista de Ciencia y Tecnología*, 8(1), 45-59.  
<https://doi.org/10.1234/rct.2018.0801.45-59>
- Lainez, A. (2019). *Tradición e innovación*. Quevedo: N°18.
- Lema Martínez, E. J. (2021). Análisis de las exportaciones de cacao tipo CCN51. *Balanza comercial*, 28.
- Litardo, F. &. (2023). *Innovaciones en la gestión de residuos agroindustriales en el cultivo del cacao*. American Environmental Association. .
- López, M. C. (7 de NOVIEMBRE de 2020).  
<https://www.redalyc.org/journal/280/28065077024/html/>
- Maldonado Mateus, J. (2023). *Herramientas Estadísticas para el Análisis de Fermentación del Cacao*. Instituto Nacional de Investigaciones Agrícolas. .

- Martínez, R. T. (2019). *Variabilidad cromática en variedades de cacao durante la fermentación*. Instituto Nacional de Investigaciones Agrícolas.
- Mathias-Rettig, A. &.-H. (2014). *Métodos para Evaluar las Características Cromáticas del Cacao*. International Food Science Association. .
- Molina. (6 de MARZO de 2019). [https://www.inia.gob.pe/wp-content/uploads/2019/12/Manual\\_Manejo\\_CacaoNativo\\_compressed.pdf](https://www.inia.gob.pe/wp-content/uploads/2019/12/Manual_Manejo_CacaoNativo_compressed.pdf)
- Nivela Andrade, M. (2020). *Modelado de la cinética de secado del cacao usando regresión lineal múltiple*. Instituto Nacional de Investigaciones Agrícolas. .
- Orbe Chamorro, M. A. (2013). *Transformación de precursores del sabor durante la fermentación y secado del cacao*. Universidad Nacional de Agricultura.
- Pérez, J. &. (2018). *Influencia de las condiciones ambientales en las propiedades cromáticas del cacao*. Universidad Nacional de Agricultura.
- Pérez, L. (2023). *Implementación de Modelos Matemáticos para la Evaluación de la Fermentación y Características Cromáticas del Cacao CCN-51*. Editorial: Universidad Nacional de Agricultura.
- Perú., M. d. (2012). *Estrategias para la sostenibilidad en la producción de cacao*. Ministerio de Agricultura de Perú.
- Prasedya, E. S. (2021). *Antioxidant and Anti-inflammatory Properties of Cocoa Bioactive Compounds*. Elsevier. .
- Quiroz, M. (22 de NOVIEMBRE de 2023). <https://repositorio.utn.edu.ec/handle/123456789/15261>
- Ruiz Aranibar, J. (2017). *Análisis de Conglomerados en la Investigación del Cacao*. Universidad Nacional de Agricultura.
- Ruiz, A. (23 de ENERO de 2020). [https://www.colombiamascompetitiva.com/wp-content/uploads/2018/10/Cadena\\_de\\_valor\\_Cacao.pdf](https://www.colombiamascompetitiva.com/wp-content/uploads/2018/10/Cadena_de_valor_Cacao.pdf)
- Sánchez Justillo, R. (2022). *Manejo integrado de plagas y enfermedades en el cultivo del cacao*. International Sustainable Agriculture Association. .
- SANDRA, O. (22 de MAYO de 2020). <https://delmonteag.com.ec/cambio-climatico-y-su-impacto-en-el-cultivo-de-cacao/>
- Smith, T. (2020). *Factores cromáticos y su impacto en la fermentación del cacao*. American Association of Food Scientists.
- Suárez, M. C. (5 de JUNIO de 2020). <https://mesioupcub.masters.upc.edu/en/xiv-summer-school-2021/courses/modelos-aditivos-generalizados-y-extensiones-aplicaciones-en-biomedicina-july-5th-to-july-9th>



- Torres, A. (2012). *Desafíos climáticos y manejo de agroquímicos en la agricultura*.  
Universidad Nacional de Agricultura.
- Uvillús Herrera, T. (2023). *Análisis de la cadena de valor del cacao fino y de aroma*. .  
International Agricultural Economics Association.
- Vasallao, M. (19 de ENERO de 2017). Cabrera, J. B. Z., Encalada, C. M., Guerrero, J. Q.,  
Reyes, S. H., Castillo, A. M., & Toro, J. L. (2021). Influencia de la madurez de las  
mazorcas de cacao: calidad nutricional y sensorial del cultivar CCN-51
- Vasallao, M. (20 de AGOTO de 2019).  
<http://revistasdigitales.utelvt.edu.ec/revista/index.php/repique/article/view/2>
- Wendy Diana Carranza Quimi. (25 de SEPTIEMBRE de 2018).  
<https://journalbusinesses.com/index.php/revista/article/view/79/288>
- Zuluaga, G. J. (12 de MAYO de 2019).  
<https://repositorio.unal.edu.co/bitstream/handle/unal/85586/52908979.2024.pdf.pdf?sequence=2&isAllowed=y>

## ANEXOS

### Código Fuente en R para el Análisis Estadístico de las Características Cromáticas del Cacao.

```
library(readxl)
library(GGally)
library(dplyr)
library(ggplot2)
library(corrplot)
library(cluster)
library(factoextra)
library(moments)
library(pheatmap)
setwd("C:/Users/user/Downloads")
cacao <- read_excel("CACAO.xlsx")

names(cacao)[3:6]<-c("Luminosidad","RG","YB","Fermentacion")

#1. Descriptivos
cacaodf<-cacao[,-2]

descrip<-cacao %>%
```

```
summarise_at(c("Luminosidad","RG","YB","Fermentacion"),list(Minimo = min,
Promedio = mean,Mediana = median, CoefAsimetria = skewness ,Maximo = max,
DesEst=sd))
```

```
clipr::write_clip(descrip)
```

```
Descriptivos<-cacaodf %>% group_by(Variedad) %>%
  summarise(across(everything(),list(Minimo = min, Promedio = mean,Mediana =
median, CoefAsimetria = skewness ,Maximo = max, DesEst=sd)))
Descriptivos
```

## #2. Diagramas de Dispersion

```
cacao %>% dplyr::select(c(Variedad,Fermentacion,Luminosidad,RG,YB)) %>%
  ggpairs(.,aes(color=Variedad, alpha=0.2))
```

```
ggplot(cacao,aes(x=Luminosidad,y=Fermentacion,colour = Variedad))+
  geom_point()+
  labs(title = "Diagrama de dispersiÃ³n Luminosidad vs Fermentacion",
  subtitle = "Autor: Wilver Santana")
```

```
ggplot(cacao,aes(x=RG,y=Fermentacion,colour = Variedad))+
  geom_point()+
  labs(title = "Diagrama de dispersiÃ³n Luminosidad Verde/Azul vs Fermentacion",
  subtitle = "Autor: Wilver Santana")
```

```
ggplot(cacao,aes(x=YB,y=Fermentacion,colour = Variedad))+
  geom_point()+
  labs(title = "Diagrama de dispersiÃ³n Luminosidad Amarillo/Azul vs Fermentacion",
  subtitle = "Autor: Wilver Santana")
```

## #3 Grafico de cajas x Variedad

```
ggplot(cacao,aes(x=Variedad,y=Luminosidad,fill = Variedad))+
  geom_boxplot()+
  coord_flip()+
  labs(title = "Diagrama de cajas Luminosidad vs Variedad",
  subtitle = "Autor: Wilver Santana")
```

```
ggplot(cacao,aes(x=Variedad,y=RG,fill = Variedad))+
  geom_boxplot()+
  coord_flip()+
  labs(title = "Diagrama de cajas Luminosidad Verde/Rojo vs Variedad",
  subtitle = "Autor: Wilver Santana")
```

```
ggplot(cacao,aes(x=Variedad,y=YB,fill = Variedad))+
  geom_boxplot()+
  coord_flip()+
  labs(title = "Diagrama de cajas Luminosidad Azul/Amarillo vs Variedad",
  subtitle = "Autor: Wilver Santana")
```

```

ggplot(cacao,aes(x=Variedad,y=Fermentacion,fill = Variedad))+
  geom_boxplot()+
  coord_flip()+
  labs(title = "Diagrama de cajas indice de Fermentacion vs Variedad",
        subtitle = "Autor: Wilver Santana")

#Matriz de Correlacion
M<-cor(cacao[,3:6])

corrplot(M, method = 'pie', insig='blank',
          addCoef.col = 'black', number.cex = 0.8,tl.cex = 0.8, order = 'FPC', diag=F,type =
"lower")

df<-as.data.frame(cacao[,3:6])

#rownames(df)<-paste0(cacao$Variedad," ",cacao$Replica)
# 4 PCA + Biplot
res.pca<-prcomp(df,scale=T)

summary(res.pca)

fviz_eig(res.pca,xlab = "Componentes",ylab="% de Varianza explicado",main =
"Gráfico de sedimentacion",subtitle="Autor: Wilver Santana"
,barfill=c("steelblue","black","steelblue","steelblue"))
fviz_contrib(res.pca, choice="var", axes = 1)+
  labs(title="Gráfico de contribuciones de las columnas \na la 1era Componente",subtitle
= "Autor: Wilver Santana"
,y="% Contribución")
fviz_contrib(res.pca, choice="var", axes = 2)+
  labs(title="Gráfico de contribuciones de las columnas \na la 2da Componente",subtitle
= "Autor: Wilver Santana"
,y="% Contribucion")
fviz_contrib(res.pca, choice="ind", axes = 1)+
  labs(title="Gráfico de contribuciones de las filas \na la 1era Componente",subtitle =
"Autor: Wilver Santana"
,y="% Contribucion")
fviz_contrib(res.pca, choice="ind", axes = 2)+
  labs(title="Gráfico de contribuciones de las filas \na la 2da Componente",subtitle =
"Autor: Wilver Santana"
,y="% Contribucion")

fviz_pca_biplot(res.pca, repel = TRUE,
  col.var = "#00AFBB", # Variables color
  col.ind = "#696969" # Individuals color,
  ,title = "Biplot del indice de Fermentacion de Cacao \nCCN-51 y sus
caracteristicasromaticas"
  ,subtitle="Autor: Wilver Santana"

```

```

    ,habillage = cacao$Variedad
    ,addEllipses = T
)

#5 Clusters + Dendograma + Heatmap
fviz_nbclust(x = df, FUNcluster = pam, method = "wss", k.max = 15,
             diss = dist(df, method = "manhattan"))+
  labs(title="Numero de clusters optimo\nMetodo: PAM", subtitle = "Autor: Wilver
Santana"
       ,y="Suma de cuadrados dentro de los clusters",x="Numero de clusters k")

set.seed(123)
pam_clusters <- pam(x = df, k = 4, metric = "manhattan")

pam_clusters

fviz_cluster(object = pam_clusters, data = df, ellipse.type = "t",
             repel = TRUE) +
  theme_bw() +
  labs(title = "Grafico de clusters\nMetodo: PAM-4 Medoids", subtitle = "Autor: Wilver
Santana") +
  theme(legend.position = "none")

fviz_nbclust(x = df, FUNcluster = kmeans, method = "wss", k.max = 15,
             diss = dist(df, method = "manhattan"))+
  labs(title="Numero de clusters optimo\nMétodo: K-means", subtitle = "Autor: Wilver
Santana"
       ,y="Suma de cuadrados dentro de los clusters",x="Numero de clusters k")

km_clusters <- kmeans(x = df, centers = 4, nstart = 50)

fviz_cluster(object = km_clusters, data = df, ellipse.type = "t",
             repel = TRUE) +
  theme_bw() +
  labs(title = "Grafico de clusters\nMétodo: k-Means, 4 Centroides", subtitle = "Autor:
Wilver Santana") +
  theme(legend.position = "none")

dfHM<-df
rownames(dfHM)<-paste0(cacao$Variedad," ",cacao$Replica)

pheatmap(scale(dfHM),treeheight_row = 0,cutree_cols = 4,fontsize = 10)

#6 Regresion Lineal Multiple
lm1<-lm(Fermentacion~.,df)

summary(lm1)

```

LME