



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería Mecánica Eléctrica

**DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO
QUE PERMITA DESCRIBIR LAS CARACTERÍSTICAS DEL FRUTO EN FUNCIÓN DEL
COLOR PARA LA PREDICCIÓN DE LA MADUREZ DEL AGUACATE HASS POST
COSECHA EN GUATEMALA**

Carlos Alberto Arellano Acevedo

Asesorado por el M.A. Jorge Mario Cabrera Madrid

Guatemala, marzo de 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO
QUE PERMITA DESCRIBIR LAS CARACTERÍSTICAS DEL FRUTO EN FUNCIÓN DEL
COLOR PARA LA PREDICCIÓN DE LA MADUREZ DEL AGUACATE HASS POST
COSECHA EN GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

CARLOS ALBERTO ARELLANO ACEVEDO

ASESORADO POR EL MTRO. ING. JORGE MARIO CABRERA MADRID

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN ELECTRÓNICA

GUATEMALA, MARZO DE 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

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

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Ing. Hugo Leonel Tiul Valenzuela
EXAMINADOR	Ing. Sergio Leonel Gómez Bravo
EXAMINADOR	Ing. Christian Antonio Orellana López
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

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

**DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO
QUE PERMITA DESCRIBIR LAS CARACTERÍSTICAS DEL FRUTO EN FUNCIÓN DEL
COLOR PARA LA PREDICCIÓN DE LA MADUREZ DEL AGUACATE HASS POST
COSECHA EN GUATEMALA**

Tema que me fuera asignado por la Dirección de la Escuela de Estudio de Posgrado, con fecha 8 de noviembre de 2022.



Carlos Alberto Arellano Acevedo



EEPFI-PP-1793-2022

Guatemala, 10 de noviembre de 2022

Director
Armando Alonso Rivera Carrillo
Escuela De Ingeniería Mecánica Eléctrica
Presente.

Estimado Ing. Rivera

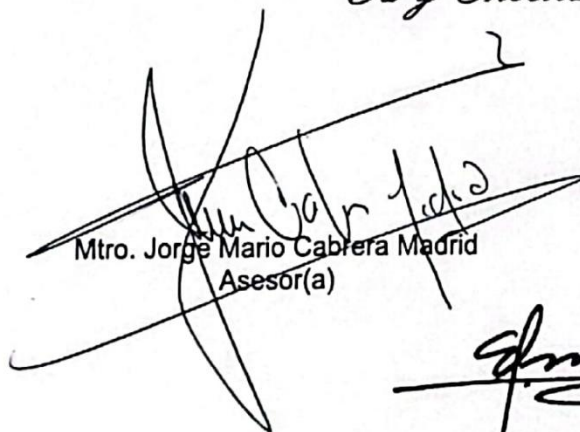
Reciba un cordial saludo de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería.


El propósito de la presente es para informarle que se ha revisado y aprobado el Diseño de Investigación titulado: **CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO QUE PERMITA DESCRIBIR LAS CARACTERÍSTICAS DEL FRUTO EN FUNCIÓN DEL COLOR PARA LA PREDICCIÓN DE LA MADUREZ DEL AGUACATE HASS POST COSECHA EN GUATEMALA**, el cual se enmarca en la línea de investigación: **Todas las áreas - Pronósticos**, presentado por el estudiante **Carlos Alberto Arellano Acevedo** carné número **201504015**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la Maestría en ARTES en Estadística Aplicada.

Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

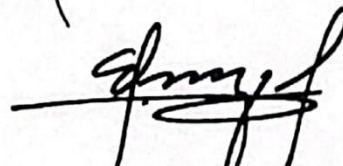
Atentamente,

"Id y Enseñad a Todos"


Mtro. Jorge Mario Cabrera Madrid
Asesor(a)


Mtro. Edwin Adalberto Bracamonte Orozco
Coordinador(a) de Maestría




Mtro. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería





EEP-EIME-1447-2022

El Director de la Escuela De Ingenieria Mecanica Electrica de la Facultad de Ingenieria de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO QUE PERMITA DESCRIBIR LAS CARACTERÍSTICAS DEL FRUTO EN FUNCIÓN DEL COLOR PARA LA PREDICCIÓN DE LA MADUREZ DEL AGUACATE HASS POST COSECHA EN GUATEMALA**, presentado por el estudiante universitario **Carlos Alberto Arellano Acevedo**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingenieria en esta modalidad.

ID Y ENSEÑAD A TODOS

Ing. Armando Alonso Rivera Carrillo
Director
Escuela De Ingenieria Mecanica Electrica

Guatemala, noviembre de 2022

LNG.DECANATO.OI.295.2023

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería Mecánica Eléctrica, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO QUE PERMITA DESCRIBIR LAS CARACTERÍSTICAS DEL FRUTO EN FUNCIÓN DEL COLOR PARA LA PREDICCIÓN DE LA MADUREZ DEL AGUACATE HASS POST COSECHA EN GUATEMALA**, presentado por: **Carlos Alberto Arellano Acevedo**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Inga. Aurelia Anabela Cordova Estrada

Decana

Guatemala, marzo de 2023

AACE/gaoc

ACTO QUE DEDICO A:

Dios	Por darme la oportunidad en la vida de atravesar este momento
Mis padres	Por su cariño incondicional e impulsarme a seguir adelante con mis metas
Mis hermanas	Quienes incondicionalmente me apoyaron brindándome consuelo y apoyo cuando mas lo necesite
Mis amigos	Por hacer de mis estudios una experiencia más divertida siendo una inspiración constante en mi vida

AGRADECIMIENTOS A:

**Universidad de San
Carlos de Guatemala**

Por permitirme realizar mis estudios creando en mi un criterio amplio

Mis amigos

Que son personas excepcionales junto a las que forme experiencias únicas e hicieron posible que cumpliera mis metas

**Mtro. Ing. Jorge Mario
Cabrerá Madrid**

Por el tiempo dedicado a su excelente asesoría para realizar este proyecto

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES	4
LISTA DE SÍMBOLOS	5
GLOSARIO	6
RESUMEN.....	7
1. INTRODUCCIÓN	8
2. ANTECEDENTES	10
3. PLANTEAMIENTO DEL PROBLEMA	14
3.1. Contexto general	14
3.2. Descripción del problema	14
3.3. Formulación del problema	15
3.4. Delimitación del problema	16
4. JUSTIFICACIÓN	18
5. OBJETIVOS	20
5.1. General.....	20
5.2. Específicos	20
6. NECESIDADES A CUBRIR Y ESQUEMA DE LA SOLUCIÓN	22
7. MARCO TEÓRICO.....	24
7.1. Estadística.....	24

7.1.1.	Tipos de estadística	24
	7.1.1.1. Estadística descriptiva.....	24
7.1.2.	Tipos de datos.....	27
	7.1.2.1. Cuantitativos y categóricos.....	28
	7.1.2.2. Discretos y continuos	28
7.1.3.	Niveles de medición	28
7.1.4.	Muestreo	29
	7.1.4.1. Muestreo aleatorio simple	29
	7.1.4.2. Muestreo sistemático	29
	7.1.4.3. Errores tipo I y tipo II	29
	7.1.4.4. Potencia de una prueba	30
	7.1.4.5. Determinación del tamaño de la muestra	30
7.1.5.	Distribución Normal	31
	7.1.5.1. Histograma	31
	7.1.5.2. Gráficas <i>QQplot</i>	32
	7.1.5.3. Pruebas de normalidad	33
7.1.6.	Modelación estadística	34
	7.1.6.1. Correlación	34
	7.1.6.2. Regresión lineal.....	35
	7.1.6.3. Análisis de varianza	36
	7.1.6.4. Pruebas Post-Hoc	38
7.2.	Color.....	39
	7.2.1. Tipos de modelos del color.....	39
	7.2.1.1. CIELAB.....	39
	7.2.1.2. RGB	40
	7.2.1.3. HSV	41
	7.2.2. Python	41
7.3.	Aguacate Hass.....	42

7.3.1.	Características del fruto	42
7.3.2.	Mediciones de Madurez.....	42
7.3.2.1.	Herramientas de medición de madurez.....	43
8.	PROPUESTA DE ÍNDICE DE CONTENIDOS	44
9.	METODOLOGÍA.....	48
9.1.	Características del estudio	48
9.2.	Unidades de análisis	49
9.3.	Operativización de variables.....	49
9.4.	Fases del estudio	50
10.	TÉCNICAS DE ANÁLISIS DE INFORMACIÓN.....	54
11.	CRONOGRAMA.....	56
12.	FACTIBILIDAD DEL ESTUDIO	58
12.1.	Recurso humano	58
12.2.	Recursos financieros	58
12.3.	Financiamiento	59
13.	REFERENCIAS.....	60
14.	APÉNDICE	66

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Gráfico de dispersión	26
2.	Ejemplo de gráfico de caja	27
3.	Distribución Normal.....	31
4.	Ejemplo de histograma	32
5.	Gráfico QQplot	33
6.	Distribución F	37
7.	Plano cartesiano con ejes RGB	40
8.	Diagrama de flujo del proceso	52
9.	Cronograma de actividades	56

TABLAS

I.	Variables del estudio.....	49
II.	Presupuesto.....	58

LISTA DE SÍMBOLOS

Símbolo	Significado
%	Porcentaje
Q	Quetzales

GLOSARIO

ANOVA	Análisis de varianza utilizado para medir la diferencia entre medias de muestras independientes categorizadas.
CIELAB	Se trata de un sistema de evaluación del color que se utiliza normalmente con mayor frecuencia para establecer las tolerancias una vez que se han establecido los límites de tono, intensidad del color y claridad.
Penetrómetro	Herramienta utilizada para medir la firmeza de un fruto.
Python	Lenguaje de programación.
Refractómetro	Instrumento utilizado para medir la concentración de sacarosa de un fruto en grados Brix.
RGB	Red, green y blue (rojo, verde y azul).

RESUMEN

La investigación se enfoca en plantear un modelo matemático para la predicción de maduración del aguacate mediante diferentes parámetros como el color y la temperatura. El problema que busca solucionar es el desperdicio del fruto y la falta de información en Guatemala a través del uso de la estadística, lo que responde a la pregunta central de proyecto.

La relevancia recae en la mejora del manejo del aguacate Hass, con la formación obtenida permitirá el planteo de automatizaciones y nuevas estrategias en el mercado.

La investigación tiene un enfoque cuantitativo con un diseño no experimental. Siendo la investigación de tipo exploratoria y descriptiva correlación con el alcance siendo correlacional transversal.

El marco teórico se enfoca en los aspectos estadísticos de la investigación, aspectos agrícolas específicos del aguacata Hass y la utilización de herramientas tecnológicas referentes al color.

1. INTRODUCCIÓN

La investigación se enfoca en poder plantear matemáticamente un modelo, de predicción para el tiempo de maduración del aguacate, mediante evaluación del color, la temperatura y los parámetros de madurez del aguacate. Es una innovación debido a que pretende descubrir algo nuevo con respecto a la tecnología utilizada para este proceso.

El problema general para tratar es el desperdicio del fruto por falta de análisis de la maduración de este, al igual que la escasez de información en Guatemala, acerca del proceso con estadísticas que permitan nuevos planteamientos.

La relevancia de la investigación recae en poder crear un manejo eficaz del aguacate Hass, permitiendo ampliar la información sobre automatizaciones y permitirá plantear, nuevas estrategias de negocios en el mercado respecto a la logística y discriminación del fruto permitiendo mayores compensaciones.

La investigación tiene un enfoque cuantitativo con un diseño no experimental (observacional), debido a la naturaleza de las variables a medir y que no se alterarán estas mismas. Siendo la investigación de tipo exploratoria y descriptiva correlacional y el alcance siendo correlacional transversal.

Se espera que la investigación muestre una correlación directa con la maduración del aguacate, que revele las mejores cámaras para el proceso y la creación de un modelo que pueda predecir la maduración respecto al color del

fruto. Esto aportará al material referencial para el uso de tecnología en métodos agrícolas y cualquier investigación subsecuente sobre la relación de estas variables.

Para plantear lo anterior, se realizarán pruebas en las cámaras para poder discriminar el tipo que es mejor para la investigación, midiéndola bajo diferentes parámetros, que puedan afectar la obtención del color. De las capturas obtenidas, se evaluarán mediante un programa desarrollado en Python, y descompuestas bajo el modelo CIEL*A*B. Siendo comparadas mediante un análisis de varianza para detectar si es igual el resultado con cada cámara. Seguido se analizará el color en función de los parámetros de madurez describiendo el fenómeno para proseguir con la modelación estadística.

La investigación propone obtener varias muestras del aguacate Hass, comparar ciertas cámaras y la utilización de herramientas para la medición de la madurez o la iluminación.

Se cuenta con todos los recursos necesarios para llevar a cabo la investigación, que será financiada por el investigador.

En el capítulo uno se hará el marco referencial, donde se describirán las investigaciones relacionadas con la investigación, que servirán como guía para los planteamientos. En el capítulo dos se describen todos los conceptos teóricos clave para el desarrollo de la investigación. En el capítulo tres se realizará la presentación de resultados. En el capítulo cuatro se hará la discusión de resultados, conclusiones y recomendaciones.

2. ANTECEDENTES

Se muestra a continuación temas que se relacionan con la investigación directa o indirectamente, con el objetivo de ampliar el conocimiento del objeto y las técnicas utilizadas para sus similares.

Respecto a los desperdicios de fruta, que puede deberse al mal manejo o la falta de predicción de este, son altos como indica Benítez (2014), donde se habla de algunas estadísticas de América Latina y el Caribe, describiendo que cerca de mil trecientos millones de toneladas de alimentos se pierden o desperdician donde las frutas, raíces, hortalizas y semillas oleaginosas, ocupan cerca de un cuarenta a cincuenta por ciento.

Para abordar el tema de la detección del color existen varias escalas, para poder medir el color de una imagen, uno de ellos es mediante un simple análisis RGB y otro es con el espacio del color CIELAB. Ambos son mediciones cuantitativas discretas del color y transformables entre sí, RGB está dividido en tres variables (R, G y B), las cuales representan la cantidad de color rojo, verde y azul.

El modelo CIELAB se divide también en tres variables (L, a y b) que representan la luminiscencia, el cambio de verde a rojo y el cambio de azul a amarillo respectivamente. Esta última escala fue diseñada para acoplarse a la visión humana, debido a esto se utilizó en una investigación similar para el análisis del aguacate con dos técnicas no destructivas (Pinto, Rueda-Chacón y Arguello, 2019). Esta parte del artículo resalta la obtención de las capturas del color por imágenes y dos de los modelos más utilizados para esta medición.

En el artículo escrito por Figueroa (2016), se menciona que se utilizó la librería de código abierto *OpenCV* y como entorno el lenguaje de programación *Python*, debido a su alta orientación al procesamiento de imágenes. En la obtención de la imagen se utilizó una cámara de resolución de 5 mega píxeles y dos puntos de iluminación, para que la reflexión de la luz máxima interviniera en la detección del color. Adicional menciona las etapas del proceso del algoritmo para trabajar la imagen. Figueroa muestra que la toma de muestras puede resultar innecesariamente compleja, por lo que se recomienda investigar si existe una manera para no depender de otros factores, además de la cámara.

Dai (2018) realizó un estudio de la piel, en el que se intenta encontrar la correlación del color obtenido, con el modelo de color CIELAB y varios compuestos del cuerpo. Menciona en el procedimiento que para medir el color de piel con el modelo CIELAB (abreviado en ocasiones como LAB donde la variable L mide la luminiscencia, a mide la cantidad de color de azul a amarillo y b la cantidad de verde a rojo), se utilizó únicamente la variable de luminiscencia del modelo "L". Con esta se realizaron los cálculos de correlación con el coeficiente de Pearson. El artículo muestra un ejemplo simple de cómo realizar una investigación con el estudio de color y como las variables que descomponen el color, pueden llegar a considerarse más relevante dependiendo del estudio dado.

En el artículo de procesamiento de imágenes de la revista de aplicaciones de computadoras (Pandey, 2013) se mencionan varias formas de modelos para obtener el color. En el caso de las manzanas, tomaron un total de treinta y nueve muestras en imágenes. Y de estas se obtuvo el color con el modelo HSI (H siendo hue o tono, S siendo saturación e I siendo intensidad). Se utilizaron otros métodos, sin embargo, en la toma de muestras se realizó un post tratamiento de la imagen que puede considerarse importante para la investigación.

Nutthatida, Jaroon y Sila (2021) desarrollaron una investigación predictiva sobre el color y los parámetros de calidad del agua. En esta se menciona el tratamiento de la muestra removiendo el fondo de la imagen y otros procesos para el manejo de la imagen. Adicionalmente se menciona que se analizaron los colores mediante la evaluación de las pocas similitudes que se pudieron evaluar usando distancias Euclidianas y de Mahalanobis. En este estudio se mencionan métodos de regresión usados en los modelos basados en una simple regresión lineal, multivariable y de mínimos cuadrados parciales. También se muestra en los modelos como se midió la diferencia de los efectos con respecto al color. El artículo es importante, ya que, ejemplifica el uso de los colores y los métodos de regresión estadística que se usaron para el estudio.

Respecto al muestreo Raggio y Moro (2008), muestran un método no paramétrico, para el muestreo de las frutas respecto a contabilización de microorganismos. En este se muestrearon tomates, duraznos, coliflor entre otros.

Crane, Douhan, Faber, Arpaia y Bender (2013), mencionó en el artículo varios parámetros postcosecha para tener en cuenta además que se menciona el muestreo que se realizó. En este se describe la medición de la madurez del fruto y como se puede medir con respecto a los grados Brix. Con el muestreo se indica la cantidad de aguacates que se tomaron de muestras, intervalos de tiempo entre cada una y tiempo.

López y Cajuste (1999), este estudio menciona la influencia que tiene la altitud de cosecha del aguacate Hass, con respecto al daño del fruto por frío. Este artículo describe los métodos usados para la medición de la maduración, así como la relevancia de la temperatura para la maduración de este.

Astudillo y Rodríguez (2018), en su estudio muestra un estudio relacionado al índice de madurez del aguacate Hass, donde el objeto fue estudiado en almacenamiento y maduración controlada. En el estudio se muestra que se utilizó una muestra de cinco frutos por cada finca y se determinó la cantidad de materia seca, para encontrar su madurez. El estudio es bastante detallado con la metodología utilizada para los parámetros de estudio.

En conclusión se inicia con un artículo, el cual ayuda a la investigación a mostrar parte de su justificación, debido a las pérdidas que hay en el mercado que luego lleva a obtener información, acerca de la cuantificación del color y con que herramientas se ha utilizado teniendo un valor amplio en el planteamiento de la metodología. Después de los artículos acerca del reconocimiento del color, inicia una recopilación de información acerca del objeto de estudio que muestran característica del fruto, pruebas y estándares de madurez a tomar en cuenta para la modelación del sistema mientras la última investigación muestra aspectos estadísticos a considerar en la investigación. En general todo esto servirá para desarrollar métodos, en base a las tecnologías y datos obtenidos por las investigaciones mencionadas antes.

3. PLANTEAMIENTO DEL PROBLEMA

3.1. Contexto general

La elección de frutas es un proceso que puede realizarse de varias formas, sin embargo, para poder automatizarlo se necesita explorar métodos, los cuales una máquina pueda analizar de maneras factibles y de manera conveniente. Al no poseer una automatización o proceso, que permita reducir el desperdicio de frutas, esto repercute en la eficiencia del proceso causando pérdidas principalmente en el vendedor y para un cliente, pues se puede incurrir en un aumento de precio en el producto. Al no haber suficientes investigaciones en el campo en Guatemala, respecto a los valores estadísticos de los parámetros de maduración del fruto, dificulta el inicio de mejoras en la eficiencia del proceso. El fruto como tal, debe tener un método que permita verificar el estado de maduración de la fruta, sin destruir el fruto mismo para averiguarlo.

3.2. Descripción del problema

En el caso del objeto de estudio, el aguacate Hass, posee ciertos parámetros que pueden afectar la madurez de este, pero se desconoce cuantitativamente los efectos que estos posean sobre él estadísticamente. En relación con el color de la fruta se sabe que indica un grado de madurez, sin embargo, no existe un modelo de predicción basado en este. Además, se desconocen varios efectos que pueda tener la maduración sobre el color, que puedan ayudar con el manejo del fruto. Si se analiza el color por medio de técnicas computacionales, también resulta lógico poder evaluar qué elementos benefician más a la toma de muestras de este parámetro y el modelo de color

que se utilizará. Otros parámetros para tomar en cuenta son: la temperatura y la humedad, debido a la ubicación de maduración de la fruta, para predecir el tiempo de vida en una determinada área geográfica, lo que involucra conocer el método de muestreo adecuado.

3.3. Formulación del problema

En este apartado se desarrollan las preguntas que definen al proyecto y forman su estructura base que marcaran las soluciones a obtener de la investigación.

- **Pregunta central**

¿Qué modelo puede predecir aceptablemente, el comportamiento de la madurez en función del color del aguacate?

- **Preguntas auxiliares**

- ¿Qué grado de correlación existe entre las características de la fruta post cosecha y el color del aguacate Hass?
- ¿Cómo se comporta el color en el aguacate Hass con respecto a sus características en el periodo de maduración post cosecha?
- ¿Qué efecto provocan en la toma del color las variables de iluminación, tipos de cámara y fondos de la imagen?

3.4. Delimitación del problema

El problema está delimitado para el área geográfica de Guatemala, siendo el objeto de estudio el aguacate de la variedad Hass, que se obtiene post cosecha.

4. JUSTIFICACIÓN

La problemática para tratar, se enmarca en el campo de la agronomía y se tratará con la línea estadística de investigación de pronósticos, debido a que se desconoce un modelo que pueda predecir la madurez de del aguacate.

La investigación será importante, para la optimización de los recursos de los involucrados en la comercialización del cultivo, permitiendo una mayor comprensión del proceso y creando la oportunidad de diseñar procesos de automatización que faciliten el manejo del aguacate Hass.

La motivación para la presente investigación es poder encontrar una mejora a los procesos selectivos del aguacate, mediante la estadística y la tecnología moderna, en este caso siendo el reconocimiento de colores mediante imágenes.

El estudio contribuirá a beneficiar los procesos de logística de la comercialización de la fruta y facilitará el proceso de automatización, que permitirá la reducción de desperdicios de la fruta e indirectamente permitirá reducir gastos de producción beneficiando a agricultores, transportistas, comercios, entre otros.

Los principales beneficiarios serán las empresas del agrícola y empresas de venta al público del fruto. Siendo indirectamente beneficiada los ingenieros que automaticen el proceso y personas con necesidad de antecedentes generados en Guatemala.

La investigación tendrá una relevancia social al aumentar la rentabilidad del fruto y permitirá motivar a la creación de nuevas automatizaciones de procesos agrícolas.

5. OBJETIVOS

5.1. General

Construir un modelo estadístico que permita describir las características del fruto en función del color para la predicción de la madurez del aguacate Hass post cosecha en Guatemala a través de la concentración total de sacarosa disuelta y la firmeza.

5.2. Específicos

1. Estimar el grado de correlación entre la fuerza de penetración de la fruta, los grados brix y el color de la fruta en diferentes tiempos, a través del coeficiente de Pearson para identificar la tendencia del modelo a construir.
2. Describir el comportamiento del color en el aguacate Hass, con respecto a los parámetros de fuerza de penetración y grados brix, mostrando datos específicos gráficamente en su etapa post cosecha para comprender la tendencia.
3. Comparar diferentes tipos de ambientes, con diferente cámara e iluminaciones para la toma de la muestra y evaluar la relevancia, con la respuesta del color utilizando un análisis de varianza y sus respectivas pruebas post hoc, para seleccionar los parámetros más adecuados del modelo.

6. NECESIDADES A CUBRIR Y ESQUEMA DE LA SOLUCIÓN

Para la presente investigación se necesita evaluar una cámara para la toma de muestras, con el fin de determinar si existe alguna diferencia entre los colores obtenidos y demostrar si es indiferente el tipo de cámara. Esto se puede comparar al medir los colores obtenidos en cada imagen de cada cámara, siendo los colores una variable cuantitativa discreta utilizando el modelo CIELAB (abreviado LAB) y tomando estos valores, para realizar un análisis de varianza entre las diferentes cámaras, que obtendrán la variable cuantitativa con los diferentes componentes de color del modelo. Adicional, se planea explorar si la iluminación puede afectar a las mismas variando esto en el proceso.

Con referencia al fruto se describirá su comportamiento, respecto al color evaluando las variables del modelo del color y su maduración, que será evaluado mediante los parámetros de materia seca y firmeza del fruto. Respecto a las variables se consideran de tipo cuantitativa continua, debido a su naturaleza de obtención y como se manejará en la investigación. Con esto se considerará realizar un análisis descriptivo, para el estado de madurez en relación con los diferentes colores. En este proceso se evaluará la normalidad que poseen los datos para poder usar técnicas paramétricas o en caso contrario su contraparte.

Para iniciar el proceso de modelaje estadístico del aguacate y su color se medirá la asociación que este posee, por un coeficiente de relación con cada componente del modelo del color CIELAB. Para el modelo se empleará el color, temperatura del ambiente, humedad y los parámetros de maduración del aguacate (fuerza de penetración en kg fuerza en el caso de la firmeza y grados brix para la medición de concentración de sacarosa) evaluando la importancia

que estas variables puedan aportar a la predicción del modelo de predicción del color del aguacate en su etapa de maduración.

7. MARCO TEÓRICO

7.1. Estadística

“Estadística es la ciencia que se encarga de planear estudios y experimentos, obtener datos y luego organizar, resumir, presentar, analizar e interpretar esos datos para obtener conclusiones basadas en ellos” (Triola, 2018, p. 2).

7.1.1. Tipos de estadística

En este apartado se definirán los tipos de estadística. Describiendo las diferencias, así como aspectos importantes y formulas que se aplican en la estadística descriptiva e inferencial.

7.1.1.1. Estadística descriptiva

Es definida como la estadística que muestra las características de los datos observados que suele ser la etapa inicial de los estudios. (González, 2007)

- Medidas de tendencia central

Estas medidas muestran los valores centrales de los datos y permiten ver el comportamiento de ellos, al igual que tomar diferentes acciones dependiendo de los valores obtenidos. En estas medidas se encuentra la media, mediana o moda. (Triola, 2018)

La media muestra el promedio de los datos y es calculada mediante la siguiente fórmula:

$$\bar{x} = \frac{\sum x}{n}$$

(Ec. 1)

Donde “x” es cada dato individual del conjunto y “n” es el número de datos.

En el caso de la mediana, a diferencia de la media, esta indica el valor intermedio de los datos y esta no cambia mucho al colocar más valores, es llamada por esto una medida de tendencia central resistente. Se calcula ordenando los datos y dependiendo de si es impar se escoge el término medio que queda y de ser par los dos elementos centrales y se realiza un promedio entre ambos.

- Medidas de dispersión

Son las medidas con las que se puede definir que, tan diferentes son entre sí los datos. Entre estas medidas se encuentra la desviación estándar, la varianza y los rangos del proceso. Siendo la desviación estándar la medida usual de medir la variabilidad, indicando la dispersión de los datos respecto a la media. Esta se denota con la letra “S” y se calcula de la siguiente forma:

$$S = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

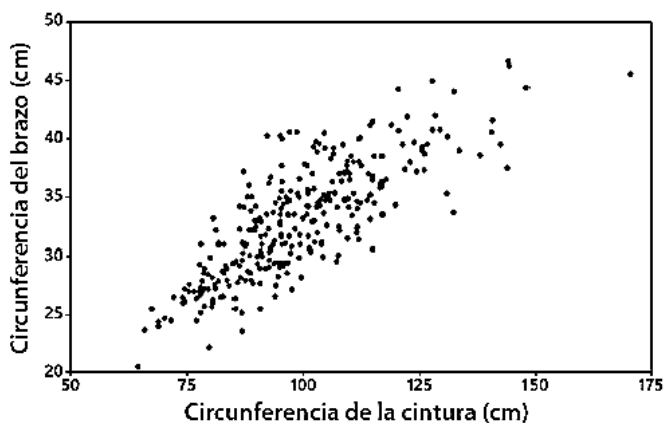
(Ec. 2)

Donde x_1, x_1, \dots, x_n son las observaciones de la muestra, n es el tamaño y \bar{x} es el promedio o media de los datos. (Gutiérrez y De la Vara, 2009)

- Diagrama de dispersión

Un diagrama de dispersión es un diagrama que, muestra los datos emparejados (x, y) , en un plano cartesiano graficando la primera variable en el eje x horizontal y la segunda en el eje vertical y . Este gráfico permite visualizar si existe una relación entre ambas variables. (Triola, 2018)

Figura 1. **Gráfico de dispersión**



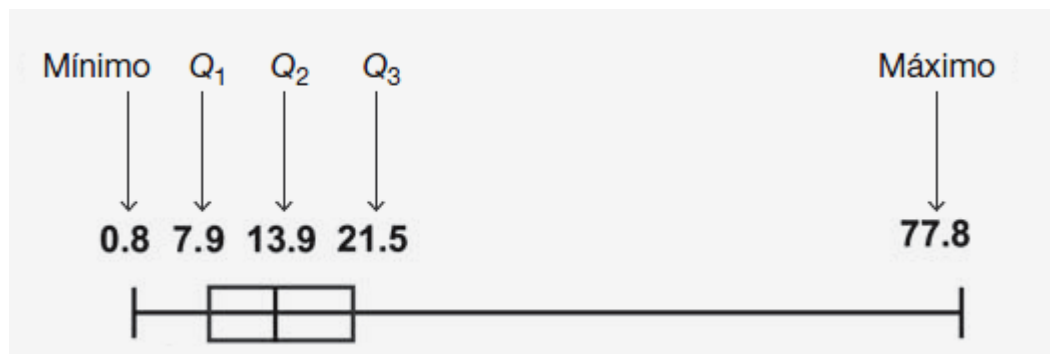
Fuente: Triola. (2018). *Estadística*.

- Gráficos de caja

Un gráfico de cajas describe el comportamiento de los datos mostrando los cuartiles (representan un cuarto de los datos cada uno siendo tres cuartiles), del proceso mostrándolo con una línea que muestra el valor mínimo, el máximo y los tres cuartiles con una línea. Debido a que no hay un estándar

como tal existe variación, en como plantear la gráfica entre programas, sin embargo, es significado de la descripción anterior se ilustra en la imagen siguiente:

Figura 2. **Ejemplo de gráfico de caja**



Fuente: Triola. (2018). *Estadística*.

- **Estadística Inferencial**

La estadística inferencial se basa en el entendimiento de las pruebas de hipótesis, es decir, en la interpretación de los valores de probabilidad para poder concluir acerca de los valores de una población (Triola, 2018).

7.1.2. Tipos de datos

En la estadística es importante diferenciar el tipo de datos, ya que, cada uno de estos posee características diversas las cuales se procesan de manera diferente lo que hace necesario identificarlas correctamente.

7.1.2.1. Cuantitativos y categóricos

Los datos categóricos o cualitativos son aquellos que permiten medir características o atributos. Mientras que los datos que se pueden contar son los datos cuantitativos. (González, 2007)

7.1.2.2. Discretos y continuos

Los datos son discretos si son valores contables finitos, es decir, si puede contarse individualmente. Al contrario, si tiene una cantidad infinita de valores y no se puede contar es del tipo continua. (Triola, 2018)

7.1.3. Niveles de medición

Esta es otra clasificación de los datos, donde puede ser: de razón, de intervalo, ordinal y nominal.

En el caso de las variables de razón, se caracterizan por coincidir con un cero natural. Mientras que, si se habla de una variable de intervalo, no posee el cero como la anterior, sino en cambio es una diferencia significativa. Estas primeras dos categorías son consideradas variables cuantitativas.

Las variables ordinales muestran que los datos tienen un orden, pero no poseen un significado como tal. Las variables nominales son aquellas que identifican y no poseen un orden como la anterior. Estas dos variables son de tipo cualitativas o categóricas. (Triola, 2018)

7.1.4. Muestreo

El muestreo se basa en tomar una parte de la población para la estimación de parámetros de ella. Este tiene diferentes técnicas que pueden considerarse como el muestreo aleatorio simple o el muestreo sistemático.

7.1.4.1. Muestreo aleatorio simple

En el muestreo aleatorio simple se selecciona una muestra de un tamaño n que tiene la misma posibilidad de ser elegida. (Scheaffer, Mendenhall y Ott, 2007)

7.1.4.2. Muestreo sistemático

El muestreo sistemático consiste en seleccionar un punto de inicio y un salto con valor k , para seleccionar la muestra y luego saltar esos k elementos para recolectar la siguiente muestra. Este muestreo es simple y evita el sesgo debido al investigador. (Scheaffer, Mendenhall y Ott, 2007)

7.1.4.3. Errores tipo I y tipo II

Al llegar a una conclusión respecto a una hipótesis, se puede presentar dos tipos importantes de error. Teniendo una hipótesis nula y una alterna, el error tipo I, se presenta cuando la hipótesis nula se rechaza cuando en realidad es verdadera. Este tipo de error es representado por *alpha* y representa esta probabilidad. Mientras que el error tipo II se presenta cuando no se rechaza la hipótesis nula cuando era falsa. En este caso suele utilizarse *beta* para representar esta probabilidad. (Triola, 2018)

7.1.4.4. Potencia de una prueba

La potencia de una prueba de hipótesis es la probabilidad complementaria de rechazar una hipótesis nula falsa. Este es calculado en base al parámetro del error tipo I, la significancia, y un parámetro poblacional. Estos datos pueden ser calculados con el software R. (Triola, 2018)

7.1.4.5. Determinación del tamaño de la muestra

Gutiérrez y De La Vara (2008) muestran que el tamaño de la muestra en caso de una distribución *t student*, que es una aproximación a la distribución normal ajustada para una cantidad menor de datos, es planteada con:

$$n = \frac{t^2_{(\alpha/2, n-1)} S^2}{E^2}$$

(Ec. 3)

Donde *t* es la transformación a la distribución *t* de la significancia α elegida por el investigador, *S* es la desviación estándar y *E* es el error de estimación que es el máximo en el que puede desviarse el valor estimado poblacional de su valor. Debido a que los grados de libertad dependen de *n*, se suele usar la aproximación de *t* igual a dos siendo la ecuación reducida la siguiente:

$$n = \frac{4S^2}{E^2}$$

(Ec. 4)

Esta ecuación puede usarse en el lenguaje de programación R con la librería *samplibgbook* la cual utiliza la función *sample size mean* para realizar el cálculo.

7.1.5. Distribución Normal

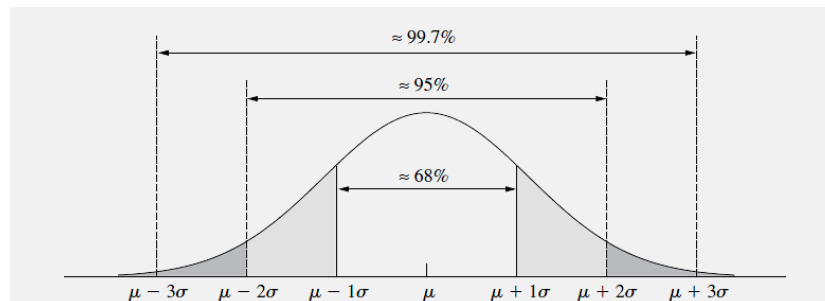
La distribución normal es uno de los modelos de poblaciones continuas, donde se puede tener cualquier valor positivo para la varianza y media. Esta sigue la siguiente función de densidad de probabilidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(Ec. 5)

Esta función es llamada curva normal. En la siguiente imagen se muestra la curva normal y el porcentaje aproximado de los datos de la población cada cierta desviación estándar de la media:

Figura 3. Distribución Normal



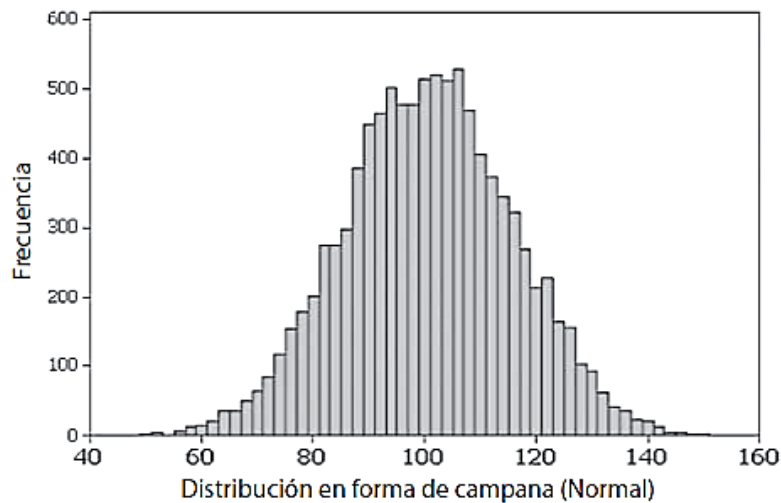
Fuente: Triola. (2018). *Estadística*.

7.1.5.1. Histograma

Un histograma muestra la distribución de los datos visualmente mediante barras que se realizan al dividir la información en grupos de ciertas frecuencias de los valores. (Gutiérrez y De la Vara, 2009)

En otras palabras, visualiza la forma de la muestra indicando las regiones donde los datos están más concentrados y donde no. (Navidi, 2006)

Figura 4. **Ejemplo de histograma**

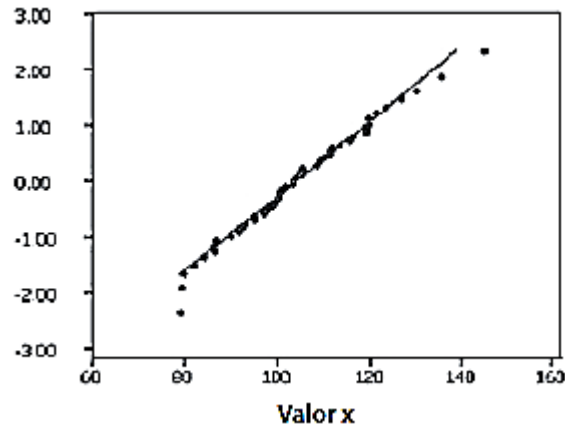


Fuente: Triola. (2018). *Estadística*.

7.1.5.2. **Gráficas *QQplot***

La gráfica *QQplot* o gráfica cuantilar normal es una gráfica cartesiana que muestra en el eje x los datos muestreados y en el eje vertical y se muestran los datos que corresponden a la distribución z. Este gráfico permite visualizar si los datos se comportan de manera normal. (Triola, 2018)

Figura 5. **Gráfico QQplot**



Fuente: Triola. (2018). *Estadística*.

7.1.5.3. Pruebas de normalidad

Además de los gráficos para visualizar si existe normalidad en los datos, es posible comprobarlo con pruebas más exactas, que los softwares modernos ya incorporan como lo son: Ji-cuadrada para bondad de ajuste, Kolmogorov, Shapiro-Wilks y Anderson-Darling. (Gutiérrez y De la Vara, 2009)

- Prueba de bondad de ajuste

Una prueba que es usada para probar la hipótesis de que los datos observados siguen cierto tipo de distribución. Si se miden las frecuencias observadas y esperadas tenemos que:

$$X^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

(Ec. 6)

Donde X^2 es el valor de una variable aleatoria, cuya distribución muestral se aproxima a chi cuadrado, con grados de libertad $v = k - 1$ y los símbolos o y e representan las frecuencias observadas y esperadas para cada elemento. La frecuencia observada siendo la cantidad de muestras multiplicado por su probabilidad. El criterio de decisión describe que no debe usarse frecuencias esperadas menores a cinco por lo que de no tenerse pueden agruparse los datos y así reducir los grados de libertad. El criterio para aceptar la hipótesis de que es equivalente a la normal es un valor pequeño de chi cuadrado. Para asegurar que se ajuste a una distribución normal se calculan las probabilidades en base a la distribución de la siguiente forma:

$$Z = \frac{x - \mu}{\sigma}$$

(Ec. 7)

Donde Z es la distribución de probabilidad normal, μ es la media, x el valor de la muestra y σ la desviación estándar. Y para el criterio de decisión se usará la distribución chi cuadrado con los grados de libertad para el valor crítico para comparar con el valor obtenido. (Walpole, Myers, Myers y Ye, 2012)

7.1.6. Modelación estadística

En esta sección se describen aspectos a tomar en cuenta para la realización y validación de un modelo estadístico, así como las estructuras posibles para su planteamiento.

7.1.6.1. Correlación

La correlación mide la relación entre dos características físicas. Uno de los coeficientes de medición numérica es el de Pearson que es denotado como

r. Este describe la medida numérica de la correlación lineal entre dos variables. Este se calcula de manera muestral y luego se estima la correlación poblacional. El valor del coeficiente esta entre menos uno y uno. El valor muestra también si los datos son directamente proporcionales o inversamente proporcionales entre sí. Dependiendo de que tan cercano a uno sea mostrara una mayor relación lineal de los datos. (Navidi, 2006)

7.1.6.2. Regresión lineal

Walpole, Myers, Myers y Ye (2012), mencionan que una forma de relacionar los datos es colocando las variables dependientes o regresores en el eje x y las respuestas en el eje y creando un modelo con la ecuación de la recta:

$$Y = \beta_0 + \beta_1 x \quad (\text{Ec. 8})$$

Donde β_0 es la intersección y β_1 es la pendiente. Si la relación es exacta es una relación determinista. De no serlo poseen una naturaleza probabilística donde el análisis de regresión lineal, permite plantear la mejor relación de esas variables y desarrolla métodos que permitan la predicción de los valores independientes. En términos simples el propósito de la regresión lineal es estimar los parámetros de la pendiente y la intersección, es decir, los coeficientes de regresión. Por lo anterior existe una estimación de la recta que se le llamará regresión ajustada. Y esta dada por la ecuación de la recta donde se calculará el valor de y pronosticado o ajustado \hat{y} , pero con los valores estimados de los coeficientes de regresión b_0 y b_1 :

$$\hat{y} = b_0 + b_1 x \quad (\text{Ec. 9})$$

La ecuación poseerá cierto error en la predicción de los valores, al determinar los coeficientes poblacionales. Para detectar el estado del modelo el concepto clave a denotar es residual. Un residual es un error en el ajuste del modelo. Siendo los n residuos grandes el ajuste del modelo no es bueno.

Para estimar los coeficientes de regresión es necesario asegurar que la suma de los mínimos cuadrados de los residuos, sea lo menor posible esto siendo conocido como el método de mínimos cuadrados.

Para la estimación de los valores de los coeficientes de regresión, se utiliza la distribución t con $n-2$. Para esto debe estimarse antes la varianza por medio del error cuadrado medio.

Para medir la calidad de ajuste se presenta el coeficiente de determinación que se denota como R cuadrado. Este coeficiente mide la proporción de la variabilidad explicada por el modelo ajustado variando entre valores de cero y uno.

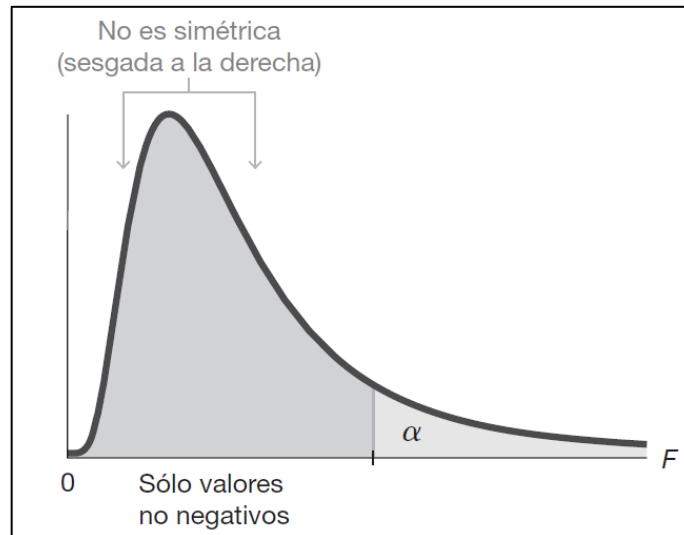
7.1.6.3. Análisis de varianza

Una prueba de análisis de varianza o ANOVA permite concluir si hay suficiente evidencia para afirmar la existencia de igualdad entre tres o más poblaciones. (Triola, 2018)

- **Distribución F**

La distribución F se caracteriza por no ser simétrica, estar sesgada a la derecha, tener solo valores positivos y de depender de dos grados de libertad. Esta distribución es usada para frecuentemente para el análisis de varianza.

Figura 6. **Distribución F**



Fuente: Triola. (2018). *Estadística*.

- ANOVA de un factor

Una prueba de análisis de varianza de un factor permite probar la igualdad entre tres o más medias poblacionales, utilizando una característica que separa los datos muestrales en diferentes categorías, es decir, un factor o tratamiento. Las condiciones para utilizarlo recaen en que la distribución de las medias sea aproximadamente normal, que posean una cantidad de tamaños de muestras iguales, que sean muestras aleatorias, que sean independientes entre sí y que las poblaciones sean categorizadas de una sola manera. Esta prueba consiste en obtener un p valor que nos permite al ser menor o igual que la confianza α rechazar la hipótesis nula la cual concluye que al menos una de las medias es diferente.

- ANOVA de dos factores

Este método como su nombre indica nos permite evaluar dos factores y tomar en cuenta la interacción entre ellos. En caso los datos segmentados por los factores se presenten diferentes a líneas paralelas se sugiere considerar su interacción. En este análisis se colocan los datos con dos divisiones en las columnas el primer factor y en las filas el segundo factor. Las categorías deben poseer el mismo tamaño de muestras, es decir, deben poseer un diseño equilibrado.

Para iniciar la prueba se debe realizar una prueba de hipótesis en base al efecto de interacción de los dos factores. Si el valor p es pequeño se rechaza la hipótesis nula de no interacción. Si se concluye esto entonces se debe detener en esta etapa de lo contrario se realiza la prueba de factor de fila que muestra si los valores de la fila son poblaciones de la misma media. Donde en este caso si el valor p es pequeño se rechaza la hipótesis nula, de que no hay efecto del factor de la fila concluyendo que hay este efecto de factor de fila existe. De igual forma se hace el mismo estudio, pero con factor de columna. (Triola, 2018)

7.1.6.4. Pruebas Post-Hoc

Estas pruebas son de utilidad cuando se rechaza que todas las medias poblacionales son iguales. En este caso estas pruebas permiten la identificación de la media que difiere en la prueba.

- Prueba de Tukey

Permite la creación de varios pares de intervalos de confianza para comparación. (Walpole, Myers, Myers y Ye, 2012)

Gutiérrez y Vara (2008) indican el siguiente calculo acerca del estadístico planteado:

$$T_{\alpha} = q_{\alpha}(k, N - k)\sqrt{CM_E/n_i}$$

(Ec. 10)

Donde CM_E es el cuadrado medio del error, n es el número de observaciones por tratamiento, k es el número de tratamientos, N-k son los grados de libertad y α representa el nivel de significancia deseado para la prueba.

7.2. Color

De acuerdo con Castro, Cerquera y Gutiérrez (2013) el color es descrito como una percepción causada por los procesos del sentido de la vista humano. En el caso estudiado es necesario medir el color, esto es posible a través de los modelos del espacio del color que permiten separar por componentes el color y obtener un dato cuantitativo que los describa.

7.2.1. Tipos de modelos del color

El color posee diferentes modelos debido a la necesidad de querer reproducirlos en diferentes medios como puede ser impreso o digital. Con esto planteado se describen algunos modelos aceptados para definir el color.

7.2.1.1. CIELAB

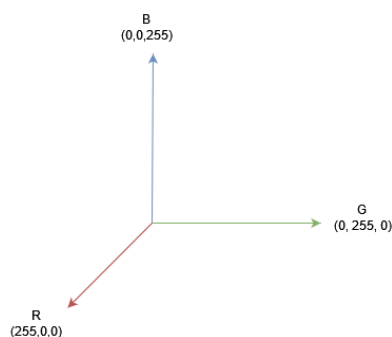
Los autores describen este modelo de manera donde descomponen el color en diferentes componentes. Para este modelo sus componentes son L, a y

b. Donde estos componentes definen la luminosidad (con mínimo de magnitud cero representando el color negro y cien como máximo representando el color blanco), como L y los colores siendo representados por las variables a y b. En el caso de a y b en su eje positivo poseen las tonalidades de rojo y amarillo, mientras en sus ejes negativo representan el color verde y azul respectivamente. Puede pensarse en estos dos últimos parámetros como colores cálidos y fríos. Según la teoría no hay máximos ni mínimos, pero en la práctica suelen tener magnitudes entre ciento veintiocho negativo a positivo.

7.2.1.2. RGB

El modelo RGB, llamado así debido a las iniciales en inglés (*red*, *green* y *blue* respectivamente), es un modelo que divide el color en tres componentes: rojo, verde y azul. Cada componente representa un plano, creando un formato de coordenadas rectangulares el cual tiene como máximo de magnitud doscientos cincuenta y cinco cada eje. Al estar los tres en su máxima magnitud forman el color blanco y en caso contrario, al estar en el centro en la posición con menor magnitud forman el color negro.

Figura 7. Plano cartesiano con ejes RGB



Fuente: elaboración propia, realizado con Microsoft Excel 365.

7.2.1.3. HSV

Aunque el modelo RGB coincide con el procesamiento biológico de los colores por la visión humana no está relacionado con la diferenciación de colores percibidos. El espacio de color HSV con sus parámetros tonalidad, saturación y valor (*Hue*, *saturation* y *value*), conforman el sistema humano de diferenciación de colores. Este espacio permite tratar por separado la información del valor del color independientemente. El inconveniente es la dificultad que posee. Este modelo este derivado del RGB para su formulación (Loesdau, Chabrier y Gabillon, 2014).

7.2.2. Python

González (2011), describe que Python es un lenguaje de programación interpretado, donde esto significa que tiene un programa intermedio llamado interprete que permite su ejecución, esto al contrario de los lenguajes compilados que se ejecutan sobre el lenguaje de máquina directamente.

Este lenguaje permite al programador realizar código complejo en pocas líneas, sin afectar la facilidad de lectura para otras personas. Comparado con otros lenguajes como C o C++, Python es lento. La ventaja es que Python puede extenderse a estos lenguajes fácilmente. Este lenguaje puede ser utilizado para el reconocimiento de imágenes y obtención del color. La librería más popular para el análisis de imágenes es *OpenCV* que aprovecha la funcionalidad del código C++ envuelta en Python para funcionar. (Mordvintsev y Abid, 2017)

7.3. Aguacate Hass

Cerdas, Montero y Cordero (2006) mencionan que el aguacate es una fruta proveniente de América central y el sur de México. Mientras que el aguacate de variedad Hass o Persea americana Mil es originaria de Guatemala.

7.3.1. Características del fruto

Entre sus características principales se encuentra que es una fruta de clima subtropical, su cultivo se da a una altitud de 1000-2000m sobre el nivel del mar, posee una resistencia intermedia al frío, con maduración tardía, superficie rugosa, cáscara negra o verde de grosor delgado, con un contenido de aceites alto, una tolerancia al almacenamiento frío alta y es un fruto climatérico, es decir, que su maduración puede ocurrir naturalmente durante el almacenamiento.

En el estudio se menciona que el rango de temperaturas de cosecha es de 16 a 18 °C como mínimo para evitar quema por escarcha.

7.3.2. Mediciones de Madurez

Astudillo y Rodríguez (2018) mencionan como parámetro fisicoquímicos del aguacate en maduración, posee bajo pruebas de sólidos solubles medidos en grados brix, contenido de sacarosa del fruto, tiene un contenido entre 5.07 y 7.26 grados brix.

7.3.2.1. Herramientas de medición de madurez

Acá se describen los diferentes instrumentos utilizados para poder capturar la información de los parámetros de madurez del aguacate Hass referentes a la dureza y grados brix.

- Refractómetro

El Refractómetro es un instrumento óptico de precisión, que se basa en refractar la luz cuando un líquido es colocado. Estos valores son definidos como el índice refractario y se relacionan con el contenido de sólidos de la muestra. Este índice es expresado en grados Brix, que es utilizado para valorar el porcentaje de sacarosa. Para obtener el valor del contenido de sólidos solubles totales se debe colocar 2 a 3 gotas del jugo en el vidrio y cerrarlo para efectuar la lectura. (Feippe y Carballo, 2003)

- Penetrómetro

El penetrómetro permite medir la firmeza de la pulpa. Este consiste en introducir el instrumento a través del fruto realizando una medición simple sin intervalos. Su medida es kg fuerza. (Feippe y Carballo, 2003)

8. PROPUESTA DE ÍNDICE DE CONTENIDOS

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTEAMIENTO DEL PROBLEMA

OBJETIVOS

RESUMEN DE MARCO METODOLÓGICO

INTRODUCCIÓN

1. MARCO REFERENCIAL

2. MARCO TEÓRICO

2.1. Estadística

2.1.1. Tipos de estadística

2.1.1.1. Estadística Descriptiva

2.1.1.2. Estadística Inferencial

2.1.2. Tipos de datos

2.1.2.1. Cuantitativos y categóricos

2.1.2.2. Discretos y continuos

2.1.3. Niveles de medición

2.1.4. Muestreo

2.1.4.1. Muestreo aleatorio simple

2.1.4.2. Muestreo sistemático

- 2.1.4.3. Errores tipo I y tipo II
 - 2.1.4.4. Potencia de una prueba
 - 2.1.4.5. Determinación del tamaño de la muestra
 - 2.1.5. Distribución Normal
 - 2.1.5.1. Histograma
 - 2.1.5.2. Gráficas *QQplot*
 - 2.1.5.3. Pruebas de normalidad
 - 2.1.6. Modelación estadística
 - 2.1.6.1. Correlación
 - 2.1.6.2. Regresión lineal
 - 2.1.6.3. Análisis de varianza
 - 2.1.6.4. Pruebas Post-Hoc
 - 2.2. Color
 - 2.2.1. Tipos de modelos del color
 - 2.2.1.1. CIELAB
 - 2.2.1.2. RGB
 - 2.2.1.3. HSV
 - 2.2.2. Python
 - 2.3. Aguacate Hass
 - 2.3.1. Características del fruto
 - 2.3.2. Mediciones de Madurez
 - 2.3.2.1. Herramientas de medición de madurez
3. PRESENTACIÓN DE RESULTADOS
4. DISCUSIÓN DE RESULTADOS

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS

ANEXOS

9. METODOLOGÍA

9.1. Características del estudio

El enfoque general de la investigación del modelo que relacione el color con su la madurez del aguacate es de tipo cuantitativo, debido a que se investigarán variables de naturaleza numéricas como en el caso de sus principales variables que son los componentes del color y los parámetros de madurez del aguacate.

El diseño, al ser una investigación de análisis del aguacate en su vida postcosecha y sus parámetros sin alterar su ambiente como tal, es de carácter no experimental (observacional) y transversal.

El alcance es exploratorio en la primera etapa de la investigación, cuando se requiere comparar las herramientas como las cámaras contra la luz evaluando su rendimiento de medidas. En la segunda parte toma un alcance descriptivo correlacional, debido a que busca relacionar el color y describir como tal el fenómeno con los parámetros de maduración del aguacate.

El alcance también es correlacional, dado que busca relacionar las características del color con los parámetros de madurez del aguacate y predecir su comportamiento entorno al periodo de vida de la fruta.

9.2. Unidades de análisis

La población en estudio estará constituida por el aguacate Hass de Guatemala, la cual se encuentra dividida en subpoblaciones dadas por unidades de aguacate en estado fisiológico similar de maduración, de la cual se extraerán muestras de forma aleatoria simple, que serán estudiadas en su totalidad.

9.3. Operativización de variables

La operacionalización de variables es un proceso que se presenta solamente en el enfoque cuantitativo debido a que las variables deben ser susceptibles a ser observadas y medidas.

Tabla I. **Variables del estudio**

Variable	Definición Teórica	Definición Operativa	Escala
Porcentaje en masa de sacarosa (%m)	Miden la concentración total de sacarosa disuelta en un líquido.	Se utilizará como uno de los parámetros para medir la madurez del aguacate al compararlo con sus valores estándar y obteniéndolo mediante el refractómetro siendo una variable dependiente medida en grados brix (°Bx)	De intervalo
Fuerza (F)	Fuerza en kg	Se utilizará un penetrómetro para medir la firmeza y reconocer la madurez del fruto siendo una variable dependiente	De razón
Iluminancia (lx)	Nivel de iluminación al momento de tomar los datos	Se utilizará un luxómetro para medir y comprobar la factibilidad de las fotos siendo una variable independiente	De razón
Temperatura (T)	Mide la temperatura del medio ambiente	Se utilizará para medir la temperatura en el periodo de tiempo de maduración de la fruta añadiendo esta variable al modelo del pronóstico siendo una variable independiente.	De intervalo

Continuación de la tabla I.

Variable	Definición Teórica	Definición Operativa	Escala
Humedad (%RH)	Mide la humedad relativa	Se utilizará para medir la humedad del ambiente siendo una variable independiente.	De intervalo
Luminosidad (L)	Cantidad de luminosidad del color definido por el modelo L*a*b	Por medio del modelo CIEL*a*b del color se medirá la luminosidad siendo el parámetro L de este valor en bits siendo una variable independiente.	De razón
Cambio de color verde a rojo (a*)	Cantidad de color verde a rojo definido por el modelo L*a*b	Por medio del modelo CIEL*a*b del color se medirá este espacio del color siendo el parámetro a de este valor	De razón
Cambio de color azul a amarillo (b*)	Cantidad de color azul a amarillo definido por el modelo L*a*b	Por medio del modelo CIEL*a*b del color se medirá este espacio del color siendo el parámetro b de este valor	De razón

Fuente: elaboración propia.

9.4. Fases del estudio

- Fase 1: revisión de literatura

La primera fase constituye la revisión de información documental y literatura tanto estadística, mediante la investigación de los modelos de regresión, sus formas de análisis y sus características como también del fruto del aguacate con sus parámetros de madurez, herramientas para medirlo y sus principales características.

- Fase 2: obtención de los materiales y comparación de cámaras

La segunda fase corresponde a la obtención de los materiales y comparación de cámaras en referente a la recolección de color en sus imágenes y definir si existe alguna diferencia entre sus capturas mediante un análisis de

varianza para definir la mejor opción para la investigación. En esta fase también se comprobará la relevancia de la iluminación en el ambiente. En caso de no detectarse cambio se procederá a escoger la cámara más económica y en caso contrario se evaluará mediante pruebas *Post hoc* las diferentes cámaras.

- Fase 3: obtención de muestras

La tercera fase consiste en la recolección de la muestra de aguacates para la captura en imagen de su color y medición de su madurez. Se utilizará una muestra de aguacate con una significancia de 0.05, un error de 0.25 y una desviación estándar de 0.7 unidades siendo una muestra de 31 unidades utilizando la ecuación cuatro del marco teórico siendo una observación directa del objeto. Todo iniciando con la obtención de la fruta en estado similar de madurez fisiológica para poder trasladar el fruto a la zona central de la ciudad para proseguir. Para esto se planea repetir el experimento con ciertas muestras para poder medir la madurez con el refractómetro y penetrómetro en diferentes tiempos de los aguacates debido a que esto solo puede realizarse una vez por la naturaleza destructiva de la prueba.

- Fase 4: análisis de los datos

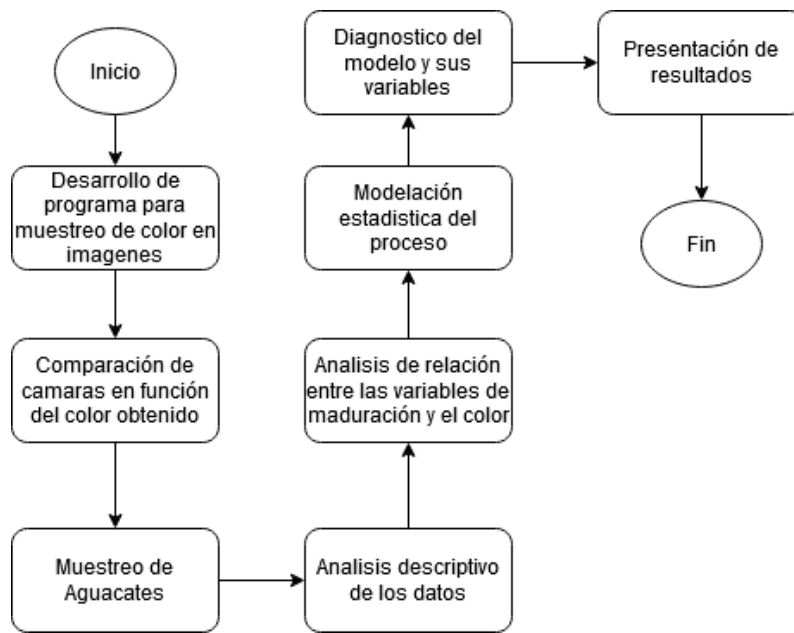
La cuarta fase consiste en el análisis de los datos obteniendo un panorama amplio del fenómeno y describiendo la conducta de las variables con diagramas de caja, dispersión e histogramas además de mostrar la correlación lineal mediante coeficiente de Pearson de cada parámetro de madurez (sólidos solubles y firmeza del fruto) y los diferentes componentes del color (L^*a^*b en el caso del modelo CIELAB).

- Fase 5: modelación estadística

La quinta fase consiste en el planteamiento del modelo estadístico y comparación con diferentes simulaciones de este para verificar su funcionalidad.

- Fase 6: redacción del informe final

Figura 8. **Diagrama de flujo del proceso**



Fuente: elaboración propia.

10. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN

A continuación, se detallarán las técnicas estadísticas que se utilizarán en el desarrollo de la investigación.

- Pruebas de Normalidad: se analizará la normalidad de los datos con diagramas *Q-Qplot*, histograma, prueba de Kolmogorov-Smirnov y pruebas de bondad de ajuste.
- Estadística descriptiva: se utilizará para medir el comportamiento del color en el aguacate utilizando diagramas de caja mostrando el análisis de tendencia central gráficamente, cálculos de medidas de tendencia central y medidas de dispersión de los datos mostrando la variabilidad.
- Pruebas de contraste: se utilizará un ANOVA de dos factores.
- Pruebas Post Hoc: se utilizará la prueba de Tukey.
- Modelación estadística: respecto al modelado estadístico de la investigación se desarrollarán pruebas sobre el modelo final de homocedasticidad, autocorrelación, multicolinealidad e independencia. Para poder comparar los distintos modelos se utilizarán los criterios de la información Akaike corregido (AIC) y bayesiano (BIC).
- Software: en el caso del software estadístico que se utilizará será R para poder realizar las gráficas descriptivas del fenómeno y poder realizar los cálculos respectivos. Y en el caso del análisis y muestreo del color se

utilizará el lenguaje de Python para la obtención de las muestras del color con la librería de reconocimiento de imágenes OpenCV.

11. CRONOGRAMA

Figura 9. Cronograma de actividades



Fuente: elaboración propia, empleando Microsoft Project 2019.

12. FACTIBILIDAD DEL ESTUDIO

12.1. Recurso humano

Para el desarrollo de la investigación es clave un profesional experto en el área de agronomía que supervise permitiendo facilitar la obtención de los recursos, entender los procesos del área necesarios para el muestreo del objeto de estudio y asistir al investigador para concluir con éxito la investigación.

El investigador profesional en el área de estadística para el correcto análisis de las variables e interpretación de estas, responsable de elaborar la investigación con los resultados obtenidos.

12.2. Recursos financieros

Se presentan los siguientes cálculos realizados por el investigador del presupuesto que será necesario para el desarrollo de la investigación.

Tabla II. Presupuesto

Recurso	Descripción	Monto Mensual	Monto Total
Humano	Investigador		
Humano	Asesor		
Tecnológico	Servicio de internet	Q 400.00	Q 2,800.00
Tecnológico	Servicio de energía eléctrica		Q 500.00
Transporte	Transporte		Q 800.00
Material	Resma de hojas		Q 500.00
Material	Instrumentos		Q 1,500.00
Material	Impresiones		Q 400.00
		Total	Q 6,500.00

Fuente: elaboración propia, empleando Microsoft Excel 365.

12.3. Financiamiento

En la sección anterior se indica el total, en quetzales, del financiamiento de la investigación siendo Q. 6,500.00 aproximadamente. Monto que será cubierto en su totalidad por el estudiante.

13. REFERENCIAS

1. Astudillo, C. y Rodríguez, P. (marzo, 2018). Parámetros fisicoquímicos del aguacate *Persea americana* Mill. cv. Hass (Lauraceae) producido en Antioquia (Colombia) para exportación. *Ciencia y Tecnología Agropecuaria*, 19(2), 383-392. Recuperado de https://doi.org/10.21930/rcta.vol19_num2_art:694..
2. Ávila, G., Sánche, E., Muñoz, E., Martínez, R. y Villalobos, E. (marzo, 2008). Diagnóstico de la calidad microbiológica de frutas y. *Revista Internacional de Botanica Experimental*, 77(1-2), 129-136.
3. Benítez, R. (22 de julio de 2014). Pérdidas y desperdicios de alimentos en América Latina y el Caribe. [Mensaje de blog]. Recuperado de <https://www.fao.org/americas/noticias/ver/en/c/239392/>.
4. Bradski, G. y Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*. Massachusetts: O'Reilly Media.
5. Castro, J., Cerquera, N. y Gutiérrez Guzmán, N. (junio, 2013). Determinación del color del exocarpio como indicador de desarrollo fisiológico y madurez en La Guayaba Pera (*Psidium guajava* cv. *Guayaba pera*), utilizando técnicas de procesamiento digital de imágenes. *Revista EIA*, 10(19), 79-89. Recuperado de <https://www.redalyc.org/articulo.oa?id=149228694007>

6. Cerdas, M., Montero, M. y Cordero Díaz, E. (2006). *Manual de Manejo Pre y Poscosecha de Aguacate (Persea americana)*. Costa Rica: Ministerio de Agricultura y Ganadería.
7. Chica-Toro, F. (junio, 2013). Grados Brix y otros parámetros de cosecha en aguacate variedad Hass (Persea americana Mill.) en el municipio de Rionegro-Antioquia. *Revista Universidad Católica de Oriente*, 35, 23-31. Obtenido de <https://revistas.uco.edu.co/index.php/uco/article/view/230>
8. Crane, J., Douhan, G., Faber, B., Arpaia, M. y Bender, G. (2013). *Cultivars and rootstocks*. Reino Unido: CAB International.
9. Dai, N.-T. (mayo, 2018). High correlation between skin color based on CIELAB color space, epidermal melanocyte ratio, and melanocyte melanin content. *peerj*, 4. doi:10.7717/peerj.4815
10. Feippe, A. y Carballo, S. (2003). *Guía práctica de análisis físico-químico de frutas y hortalizas*. Canelones: INIA Las Brujas.
11. Figueroa, E. (junio, 2016). Sistema de visión artificial para la identificación del estado de madurez de frutas (granadilla). *Redes de Ingeniería*, 7(1), 78–86. Doi:<https://doi.org/10.14483/udistrital.jour.redes.2016.1.a08>.
12. González, L. (2007). *Estadística Descriptiva y Probabilidad*. Bogotá: Universidad de Bogotá Jorge Tadeo Lozano.

13. González, R. (2011). *Python Para Todos*. España: Creative Commons Reconocimiento 2.5. Recuperado de <https://duenaslerin.com/tico2/pdfs/python-para-todos.pdf>.
14. Gutiérrez, H. y De La Vara, R. (2008). *Análisis y diseño de experimentos*. México D.F.: McGraw-Hill.
15. Gutiérrez, H. y De la Vara, R. (2009). *Control estadístico de calidad y seis sigma*. México, D.F.: McGrawhill.
16. Loesdau, M., Chabrier, S. y Gabillon, A. (julio, 2014). Hue and Saturation in the RGB Color Space. *Lecture Notes in Computer Science*, 8509. Doi:https://doi.org/10.1007/978-3-319-07998-1_23
17. López, L. Cajuste, J. (1999). Comportamiento postcosecha de fruta de aguacate cv. Hass con base en la altitud de producción y tipo de floración. *Revista Chapingo Serie Horticultura*, 5, 365-371. Recuperado de http://www.avocadosource.com/Journals/CICTAMEX/CICTAMEX_1997/Agr_1_97.pdf.
18. Mordvintsev, A. y Abid, K. (2017). *OpenCV-Python Tutorials*. USA: OpenCV-Python.
19. Navidi, W. (2006). *Estadística para Ingenieros y Científicos*. México: McGraw-Hill Companies.

20. Nutthatida P., Jaroon, J. y Sila K. (abril, 2021). Investigation into the predictive performance of colorimetric sensor strips using RGB, CMYK, HSV, and CIELAB coupled with various data preprocessing methods: a case study on an analysis of water quality parameters. *Journal of Analytical Science and Technology* 12(19), 1-16. <https://doi.org/10.1186/s40543-021-00271-9>
21. Pandey, R. (2013). Image Processing and Machine Learning for Automated. *International Journal of Computer Applications*, 81(16), 32-33. Recuperado de <https://www.ijcaonline.org/archives/volume81/number16/14209-2455>.
22. Phuangsaikai, N., Jakmunee, J. y Ki, S. (abril, 2021). Investigation into the predictive performance of colorimetric sensor strips using RGB, CMYK, HSV, and CIELAB coupled with various data preprocessing methods: a case study on an analysis of water quality parameters. *Journal of Analytical Science and Technology*, 12(19). doi:<https://doi.org/10.1186/s40543-021-00271-9>
23. Pinto, J., Rueda-Chacón, H. y Arguello, H. (agosto, 2019). Classification of Hass avocado (persea americana mill) in terms of its ripening via hyperspectral images. *TecnoLógicas*, 22(45), 109-128. doi:<https://doi.org/10.22430/22565337.1232>
24. Scheaffer, R., Mendenhall, W. y Ott, L. (2007). *Elementos de muestreo*. España: Paraninfo, S.A.

25. Triola, M. (2018). *Estadística*. Ciudad de México: Pearson Educación de México, S.A. de C.V.
26. Walpole, R., Myers, R., Myers, S. y Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias*. México: Pearson.

14. APÉNDICE

Apéndice 1. Matriz de coherencia

Problema estadístico	Preguntas de investigación	Objetivos	Metodología
Se desconoce un modelo de predicción para la maduración del aguacate Hass post cosecha en Guatemala	¿Qué modelo puede predecir aceptablemente, el comportamiento de la madurez en función del color del aguacate?	Construir un modelo estadístico que permita describir las características del fruto en función del color para la predicción de la madurez del aguacate Hass post cosecha en Guatemala a través de la concentración total de sacarosa disuelta y la firmeza.	El enfoque general de la investigación del modelo que relacione el color con su la madurez del aguacate es de tipo cuantitativo, debido a que se investigaran variables de naturaleza numéricas como en el caso de sus principales variables que son los componentes del color y los parámetros de madurez del aguacate.
01. No se conoce la asociación que posee el color con las características de la fruta post cosecha del aguacate Hass.	01. ¿Qué grado de correlación existe entre las características de la fruta post cosecha y el color del aguacate Hass?	01. Estimar el grado de correlación entre la fuerza de penetración de la fruta, los grados brix y el color de la fruta en diferentes tiempos a través del coeficiente de Pearson para identificar la tendencia del modelo a construir.	El diseño, al ser una investigación de análisis del aguacate en su vida postcosecha y sus parámetros sin alterar su ambiente como tal, es de carácter no experimental (observacional).
02. Se desconoce el comportamiento de los colores (modelo L^*a^*b o HCL) para un aguacate Hass post cosecha con relación a su característica post cosecha	02. ¿Cómo se comporta el color en el aguacate Hass con respecto a sus características en el periodo de maduración post cosecha?	02. Describir el comportamiento del color en el aguacate Hass con respecto a los parámetros de fuerza de penetración y grados brix mostrando datos específicos gráficamente en su etapa post cosecha para comprender la tendencia.	El tipo de estudio es exploratorio en la primera etapa de la investigación cuando se requiere comparar las herramientas como las cámaras contra la luz evaluando su rendimiento de medidas. En la segunda parte toma un tipo de estudio descriptivo correlacional, debido a que busca relacionar el color y describir como tal el fenómeno con los parámetros de maduración del aguacate.
03. Se desconoce si la iluminación o cámara puede interferir o mejorar en la toma de muestras de color	03. ¿Qué efecto provocan en la toma del color las variables de iluminación, tipos de cámara y fondos de la imagen?	03. Comparar diferentes tipos de ambientes con diferente cámara e iluminaciones para la toma de la muestra y evaluar la relevancia con la respuesta del color utilizando un análisis de varianza y sus respectivas pruebas post hoc para seleccionar los parámetros más adecuados del modelo.	El alcance es de tipo descriptivo correlacional transversal, dado que busca relacionar las características del color con los parámetros de madurez del aguacate y predecir su comportamiento entorno al periodo de vida de la fruta.

Fuente: elaboración propia.