



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

**DISEÑO DE UN MÉTODO DE MUESTREO PARA LA MEDICIÓN DE LA CALIDAD DE
RACIMOS DE PALMA DE ACEITE EN ALTA VERAPAZ, GUATEMALA**

Inga. Ana Sofía Rogel Palencia

Asesorado por Mtro. Edwin Adalberto Bracamonte Orozco

Guatemala, octubre de 2021

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE UN MÉTODO DE MUESTREO PARA LA MEDICIÓN DE LA CALIDAD DE
RACIMOS DE PALMA DE ACEITE EN ALTA VERAPAZ, GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA

POR

INGA. ANA SOFÍA ROGEL PALENCIA

ASESORADO POR MTRO. EDWIN ADALBERTO BRACAMONTE OROZCO

AL CONFERÍRSELE EL TÍTULO DE

MAESTRA EN ESTADÍSTICA APLICADA

GUATEMALA, OCTUBRE DE 2021

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

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

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Mtr. Ing. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtr. Ing. Walter Arnoldo Bardales Espinoza
EXAMINADOR	Mtr. Ing. Edgar Dario Álvarez Coti
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 UN MÉTODO DE MUESTREO PARA LA MEDICIÓN DE LA CALIDAD DE RACIMOS DE PALMA DE ACEITE EN ALTA VERAPAZ, GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 24 de noviembre de 2018.

Inga. Ana Sofía Rogel Palencia



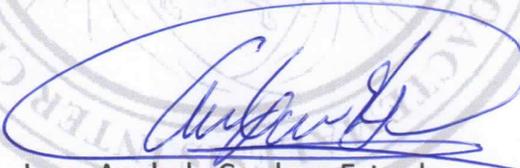
USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

Decanato
Facultad de Ingeniería
24189101 - 24189102

DTG. 534.2021

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Postgrado, al Trabajo de Graduación titulado: **DISEÑO DE UN MÉTODO DE MUESTREO PARA LA MEDICIÓN DE LA CALIDAD DE RACIMOS DE PALMA DE ACEITE EN ALTA VERAPAZ, GUATEMALA**, presentado por la **Ingeniera Ana Sofía Rogel Palencia**, estudiante del programa de **Maestría en Artes en Estadística Aplicada**, y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:


Inga. Anabela Cordova Estrada
Decana



Guatemala, octubre de 2021

AACE/cc



Guatemala, octubre de 2021

LNG.EEP.OI.035.2021

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

“DISEÑO DE UN MÉTODO DE MUESTREO PARA LA MEDICIÓN DE LA CALIDAD DE RACIMOS DE PALMA DE ACEITE EN ALTA VERAPAZ, GUATEMALA”

presentado por **Ana Sofía Rogel Palencia** quien se identifica con carné **201021769** correspondiente al programa de **Maestría en artes en Estadística aplicada**; apruebo y autorizo el mismo.

Atentamente,

“Id y Enseñad a Todos”

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



Guatemala, 31 de octubre de 2020

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

Estimado Señor Director:

De manera atenta hago constar que he revisado el **Trabajo de Graduación** de la estudiante **Ana Sofía Rogel Palencia**, de la Maestría en Artes en Estadística Aplicada, titulado **“Diseño de un método de muestreo para mejorar la medición de la calidad de racimos de palma de aceite en alta Verapaz, Guatemala”**.

Con base en la evaluación realizada hago constar la originalidad, calidad, coherencia según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobados por la Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Cumpliendo tanto en su estructura como en su contenido, por lo cual el trabajo evaluado cuenta con mi aprobación.

Atentamente,

“Id y Enseñad a Todos”



Mtro. Edwin Adalberto Bracamonte Orozco
Coordinador

Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado
Facultad de Ingeniería

Guatemala, 31 de octubre de 2020

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

Estimado Señor Director:

De manera atenta hago constar que he revisado el **Trabajo de Graduación** de la estudiante **Ana Sofía Rogel Palencia**, de la Maestría en Artes en Estadística Aplicada, titulado **“Diseño de un método de muestreo para mejorar la medición de la calidad de racimos de palma de aceite en alta Verapaz, Guatemala”**.

Con base en la evaluación realizada hago constar la originalidad, calidad, coherencia según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobados por la Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Cumpliendo tanto en su estructura como en su contenido, por lo cual el trabajo evaluado cuenta con mi aprobación.

Atentamente,

“Id y Enseñad a Todos”



Mtro. Edwin Adalberto Bracamonte Orozco
Asesor

Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado
Facultad de Ingeniería

ACTO QUE DEDICO A:

Mis padres	Guillermo Rogel y Mayra Palencia por ser fuente infinita de inspiración y superación.
Javier Álvarez	“Qué lindo que estás, sos un caramelo”. Gracias por tu apoyo incondicional.
Familia Rogel Álvarez	Por ser parte de mi corazón y de mis logros.
Familia Rogel Batz	Hermanos de mi corazón y compañeros de aventuras.
Fernando Itzep	Por cada aventura vivida y su valiosa amistad.
Jorge Mario Corzo	Por todas las enseñanzas que me trajeron hasta aquí. Le guardo un profundo afecto
Mis amigos	Flory Jiménez, Daisy Rosales, Maritza Axjollip, Darwin Jocholá, Tracy Vela, Marvin Pec, Andrea Guerra y todos y cada uno de ustedes que compartieron esta experiencia junto a mí.

Edwin Bracamonte y Por su importante asesoría y cariño.
Mayra Castillo

AGRADECIMIENTOS A:

**Universidad de San
Carlos de Guatemala**

Por ser *alma máter* de conocimiento.

Facultad de Ingeniería

Por ser casa de Ciencia y aprendizaje.

**Edwin Bracamonte y
Mayra Castillo**

Por el empeño y dedicación que siempre concedieron a nuestra formación.

Roberto Hernández

Por su invaluable apoyo y guía.

**Mis amigos de Fray
Bartolomé de las Casas**

Andrés Pitan, Bartolomé Chun, Edgar Delgado, Juan Bernabé, Isaias Álvarez, Leonel Bolom, Juan García, Ronald Cacao, Alejandro Cornelio y José Girón por su apoyo incondicional en gestión de información para la conformación del presente documento.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	III
LISTA DE SÍMBOLOS	V
GLOSARIO	VII
RESUMEN.....	IX
PLANTEAMIENTO DEL PROBLEMA.....	XI
OBJETIVOS.....	XV
RESUMEN DEL MARCO METODOLÓGICO	XVII
INTRODUCCIÓN	XXV
1. MARCO REFERENCIAL.....	1
1.1. Estudios Previos.....	1
1.2. Marco contextual	5
2. MARCO TEÓRICO.....	9
2.1. Palma de aceite.....	9
2.1.1. Taxonómica y morfología	11
2.2. Medición de calidad de racimos	12
2.3. Fundamentos Estadísticos	16
2.4. Estadística Inferencial.....	16
2.4.1. Probabilidad.....	17
2.4.2. Espacio muestral	18
2.5. Muestreo estadístico	19
2.5.1. Tipos de muestreo	21
2.6. Distribución de probabilidad	23
2.6.1. Distribución de frecuencias.....	24

2.7.	Histograma.....	24
2.8.	Distribuciones para variables discretas.....	25
2.9.	Distribución para variables continuas.....	26
2.10.	Distribuciones muestrales.....	28
2.11.	Pruebas de bondad de ajuste.....	31
2.11.1.	Prueba X^2	32
2.11.2.	Prueba de Kolmogorov-Smirnov.....	33
2.11.3.	Jarque-Bera.....	35
2.12.	Métodos paramétricos para el contraste de hipótesis.....	37
2.13.	Prueba de Grubbs.....	39
3.	PRESENTACIÓN DE RESULTADOS.....	41
3.1.	Distribución de probabilidad de ajuste al peso de racimos de palma de aceite.....	41
3.2.	Tamaño muestral requerido para la medición de la calidad de fruta.....	46
3.3.	Comparación de dos métodos de muestreo para la selección de la muestra de racimos.....	50
4.	DISCUSIÓN DE RESULTADOS.....	53
	CONCLUSIONES.....	59
	RECOMENDACIONES.....	61
	REFERENCIAS.....	63
	APÉNDICES.....	69
	ANEXOS.....	77

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Peso de la población de racimos en centro de acopio	XIX
2.	Racimos en planta de beneficio a) rampa; b) peso de la muestra.	XXII
3.	Método de muestreo a) cuerda; b) cuadrante	XXIII
4.	Distribución del cultivo y plantas de beneficio en Guatemala.....	7
5.	Área cultivada en América Latina (2000 -2014)	10
6.	Categorización de fruta a) verde; b) maduro; c) sobre maduro.....	15
7.	Espacio muestral de eventos excluyentes y no excluyentes.....	18
8.	Histograma.....	25
9.	Distribución a) Binomial; b) Poisson; c) Hipergeométrica.....	26
10.	Distribución a) t <i>Student</i> ; b) X^2 ; c) F-Snedecor; d) Gamma.....	27
11.	Curva de distribución normal.....	28
12.	Gráfico de predichos y residuales.	35
13.	Área de aceptación y rechazo para H_0 (α 5 %).....	38
14.	qqplot medición 13, identificación de puntos atípicos	45
15.	Tamaño muestral y la desviación estándar poblacional.....	47
16.	Correlación entre la media poblacional y las medias de muestreo	49
17.	Intervalo para el error de estimación.....	50
18.	Correlación de medias para métodos de muestreo.....	52

TABLAS

I.	Descripción de variables e indicadores de los objetivos	XVIII
II.	Descripción de lotes para peso de racimos	XX
III.	Clasificación de racimos basada en frutos sueltos y color	14
IV.	Clasificación de madurez de racimos con base en fruto suelto	14
V.	Número de racimos por medición	41
VI.	Valores críticos para prueba de Jarque-Bera	43
VII.	λ para transformación y p-valor para bondad de ajuste.	44
VIII.	Tamaño de muestra por medición	46
IX.	Comparación de coeficientes de variación	48
X.	Comparación de medias para métodos de selección de muestra	51

LISTA DE SÍMBOLOS

Símbolo	Significado
S^2	Desviación estándar de la muestra
σ	Desviación estándar de la población
$^{\circ}\text{C}$	Grados Celsius
kg	Kilogramo
B	Límite para el error de estimación
>	Mayor que
\bar{y}	Media de la muestra
μ	Media de la población
m_4	Medida de curtosis
<	Menor que
m	Metro
mm	Milímetro
%	Porcentaje
α	Significancia estadística
Σ	Sumatoria
TM	Tonelada métrica
m_3	Valor de asimetría
S^2	Varianza de la muestra
σ^2	Varianza de la población

GLOSARIO

Beneficio	Término utilizado como referencia al proceso de extracción de aceite de los racimos de palma aceitera.
bmh- S (c)	Bosque muy Húmedo subtropical cálido.
FAO	Organización de las Naciones Unidas para la Alimentación y la Agricultura.
Mesocarpio	Capa intermedia del pericarpio de los frutos, que se encuentra entre el endocarpio y el epicarpio.
Método de la Cuerda	Forma de selección de una muestra de racimos a través de una cuerda con nudos equidistantes.
Método del cuadrante	Método propuesto para la selección de muestra en el que se dibuja un plano cartesiano imaginario sobre el contenido de racimos de una unidad de transporte y se toma una muestra compuesta en los cuatro cuadrantes resultantes.
Nivel de confianza	Probabilidad de que el parámetro poblacional se encuentre entre el intervalo de confianza del estimador.

Racimo	Conjunto de frutos que se desarrollan en un mismo tallo.
Significancia	Probabilidad de caer en el error tipo I, rechazar la hipótesis nula cuando es verdadera.

RESUMEN

El cultivo de palma de aceite se encuentra distribuido en la región norte, nororiente y sur de Guatemala. Debido a su importancia y a la carencia de información específica para las condiciones de producción del país se propuso validar la metodología de muestreo para la medición de calidad de fruta.

Para ello se comprobó la normalidad del peso de racimos, paso que permitió fundamentar los análisis siguientes, bajo los procedimientos de la estadística paramétrica. Los resultados especifican el tamaño de muestra requerido para estimaciones representativas y significativas al 5 % y también se incluye en la discusión, la comparación de dos métodos para la selección de muestra.

La metodología se trabajó con base en estudios previos, realizados en Malasia y Colombia, principalmente los estudios efectuados por García y Yáñez (2000), quienes sugirieron que el peso de racimos se distribuye de manera normal y que para realizar el muestreo deben seleccionarse de forma aleatoria 30 racimos por medio del método de la cuerda.

Los hallazgos del presente estudio sugieren que el peso de las poblaciones de racimos no siempre se comporta de forma normal y que una fuente de variación es el desprendimiento de fruto. El número de racimos requerido para el muestreo es 27 y tanto el método de selección de la cuerda como el del cuadrante son representativos, de tal forma que la metodología propuesta por García y Yáñez (2000) es válida para la medición de calidad de fruta para plantas de beneficio en Guatemala.

PLANTEAMIENTO DEL PROBLEMA

- Contexto general

Las empresas que se dedican a la extracción de palma de aceite (*Elaeis guineensis Jacq*), se encuentran en La Franja Transversal del Norte, Izabal, Petén y parte de la Costa Sur de Guatemala, según la Gremial de Palmicultores de Guatemala (GREPALMA, 2017).

Las plantas de beneficio de aceite generalmente se ubican a una distancia estratégica que favorezca la logística de traslado ya que, tal como lo expresan Erales (2016) y Taillez, Siaka, Bonny y Jacquemard (1996) la calidad del aceite producido depende directamente de la calidad del fruto proveniente de las plantaciones, por ello, parte de las actividades y recursos de la industria son dirigidas a la medición de la calidad de racimos.

La metodología de evaluación que ha sido empleada desde el año 2003 fue seleccionada con base a lo descrito por García y Yáñez (2000) en sus estudios realizados para determinar una metodología alterna para análisis de racimos y muestreo de racimos en tolva. En estos estudios se describe que, para obtener una muestra representativa y significativa, son necesarios 30 racimos seleccionados aleatoriamente por medio del método de la cuerda, todo ello, para las condiciones de la zona palmera Norte y Occidental de Colombia.

La metodología anteriormente descrita es la que se utiliza en Guatemala, se aplica a camiones con una capacidad de carga de 14 a 24 TM, indistintamente de las condiciones del cultivo. Algunos factores como el año de siembra, variedad

del cultivo, condiciones climáticas, ciclos de cosecha, logística de recepción, etc. pueden alterar la variabilidad y la metodología propuesta en Colombia podría no ser válida para las condiciones de otra región o país, razón por la cual vale la pena generar información para las condiciones de Guatemala, específicamente en la zona de interés para el presente estudio, Fray Bartolomé de las Casas, Alta Verapaz.

- Descripción del Problema

El objetivo de la inferencia estadística radica en obtener estimadores que describan los parámetros de las poblaciones de una forma confiable y representativa. Para conocer el tamaño muestral, el método de muestreo y los estimadores de parámetros de una población de racimos contenidos en una unidad de transporte, es necesario hacer uso de las técnicas de muestreo probabilístico. El muestreo de racimos en rampa de recepción de fruta es una actividad importante debido a que el grado de madurez se encuentra directamente relacionado con la calidad del aceite producido y comercializado al consumidor final.

Para cumplir las actividades de evaluación de racimos y generar información para la palmicultura guatemalteca se requiere emplear metodologías ajustadas a las condiciones de Guatemala. Es necesario conocer la distribución de frecuencias a la que se ajusta la población de racimos, determinar el tamaño muestral, evaluar los métodos de muestreo y comparar el método actualmente utilizado, con el método propuesto en el estudio.

No existe en Guatemala un método de muestreo definido para el muestreo de racimos, las metodologías conocidas y utilizadas provienen de estudios en Colombia, en dónde la distribución de probabilidad de los pesos de racimos se

ajusta a la Normal, donde los registros climáticos reportan temperaturas medias de 28 – 32 °C y precipitaciones entre 1129 - 3000 mm al año, en plantaciones de 10 a 13 años y provenientes de material genético Papúa (García y Yáñez, 2009).

De acuerdo con la información de estaciones meteorológicas provista por empresas de palma ubicadas en la Fray Bartolomé de las Casas, la temperatura media y la precipitación anual de las zonas de producción oscilan entre 25.4 - 25.6 °C y 2200 - 3700 mm. La fruta procesada en la planta de beneficio proviene de plantaciones de 4 a 19 años y de 21 materiales genéticos. Por lo tanto, resalta la conveniencia de comprobar el comportamiento normal de las poblaciones de racimos bajo condiciones de producción en Guatemala.

- Formulación del problema

Para estructurar la formulación del problema se enunciaron una pregunta central y tres auxiliares, estas preguntas tienen relación directa con los objetivos en secciones posteriores.

- Pregunta central

¿Cuál es la metodología de muestreo estadísticamente significativa y representativa para evaluar la calidad de fruta ingresada a una planta de beneficio por unidad de transporte, de acuerdo con las condiciones de producción de Fray Bartolomé de las casas?

- Preguntas auxiliares
 - ¿A qué distribución de probabilidad se ajustan las medidas del peso de racimos en la zona de producción de Fray Bartolomé de las Casas?
 - ¿Cuál es el tamaño muestral requerido para un muestreo representativo y significativo para la medición de la calidad de los racimos?
 - ¿Qué método de muestreo es el más práctico y el más representativo para el muestreo de racimos?

- Delimitación del problema

El estudio se realizó durante el año 2018 y 2019, en las instalaciones para recepción de fruta de una planta extractora, ubicada en el municipio de Fray Bartolomé de las Casas, Alta Verapaz, Guatemala.

La unidad de análisis para las evaluaciones fueron los racimos contenidos en los medios de transporte que utilizan las fincas productoras para el traslado de los racimos a planta de beneficio.

El enfoque que se utilizó para el análisis de la información fue eminentemente estadístico, bajo los lineamientos de muestreo probabilístico.

OBJETIVOS

General

Diseñar un método de muestreo para la medición de calidad de fruta de palma de aceite (*Elaeis guineensis Jacq*) ingresada a planta de beneficio por medio de estimaciones basadas en la distribución probabilística del peso de los racimos para mejorar el muestreo realizado y ajustarlo a las condiciones de Fray Bartolomé de las Casas, Alta Verapaz, Guatemala.

Específicos

- Determinar la distribución de probabilidad a la que se ajusta el peso de racimos para seleccionar métodos de muestreo válidos por medio de pruebas de bondad de ajuste.
- Determinar el tamaño muestral requerido para la medición de la calidad de fruta por medio de métodos ajustados a la distribución de probabilidad.
- Comparar por medio de pruebas de contraste de hipótesis, dos métodos de selección de la muestra para sustentar la representatividad y significancia de los métodos propuestos por el estudio.

RESUMEN DEL MARCO METODOLÓGICO

El estudio posee un enfoque cuantitativo con variables continuas, en donde se aplicaron conceptos de muestreo estadístico para hacer inferencias sobre la población de racimos y mejorar el proceso de muestreo de calidad de fruta en plantas de beneficio.

El diseño del estudio es no experimental, ya que los métodos bajo los cuales se realizaron las mediciones se basan en los conceptos del muestreo estadístico y prueba de hipótesis.

El estudio es considerado de alcance descriptivo correlacional, ya que describe los estimadores poblacionales y el comportamiento en la distribución de frecuencias del peso de racimos para realizar inferencias.

- Variables e indicadores

Las variables e indicadores que describen la manera en que se trabajaron los objetivos propuestos se describen en la tabla I.

Tabla I. **Descripción de variables e indicadores de los objetivos**

Variable	Definición	Indicador
Peso (kg)	Medida de fuerza gravitatoria que actúa sobre un objeto.	Pruebas de bondad de ajuste para la distribuciones de probabilidad Normal con $\alpha = 5 \%$.
Número de racimos	Enumeración de racimos.	$n = \frac{N\sigma^2}{(N - 1)D + \sigma^2}$

Fuente: elaboración propia.

- Fases de la investigación
 - Fase 1. Revisión bibliográfica

Para el diseño del estudio y para el análisis e interpretación de resultados se recopilaron conceptos estadísticos que conforman el fundamento dentro del marco de la presente investigación científica. La revisión bibliográfica se enfocó al desarrollo de conceptos tales como la morfología de la palma de aceite, medición de calidad de racimos en planta de beneficio, distribuciones de probabilidad y muestreo estadístico.

- Fase 2. Gestión y recolección de la información
 - Medición de peso de racimos

Se realizaron 15 pesos de racimos en plantaciones de palma de aceite ubicadas en el municipio de Fray Bartolomé de las Casas, Alta Verapaz. Cada peso de racimos fue enfocado a un centro de acopio de fruta en campo, cada centro de acopio contiene los racimos que serán destinados a una unidad de transporte. El personal encargado de la actividad de peso de racimos registró el número y peso de cada racimo contenido en los centros de acopio y luego los acomodaron en la unidad de transporte, esto último, para asegurar el pesaje de todos los racimos contenidos en la unidad de transporte (figura 1). El peso se determinó por medio de una balanza romana de resorte y las medidas fueron registradas en kilogramos (kg).

Figura 1. **Peso de la población de racimos en centro de acopio**



Fuente: elaboración propia.

El formato utilizado para el registro de información se detalla en el apéndice 1. Los pesos de racimos fueron realizados en los meses de junio a diciembre del año 2018. Los lotes seleccionados para muestreo se tomaron de forma aleatoria, dentro del grupo de lotes programados a cosechar. Solamente se seleccionaron los lotes que tuvieran ≥ 10 días y ≤ 15 días entre la cosecha anterior y la cosecha del día de la medición, para evitar sesgos en la información debidos al desprendimiento de fruto.

Tal como se describe en la tabla II se tomaron en cuenta, para la realización del muestreo cinco fincas, cuatro variedades y cinco proyectos (el proyecto hace referencia al año de siembra).

Tabla II. **Descripción de lotes para peso de racimos**

Medición	Finca	Lote	Variedad	Proyecto
1	1	8	Deli X Ghana	2008
		9	Deli X Ghana	2008
2	1	3	Deli X La mé	2008
		4	Deli X La mé	2008
3	2	2	Deli X Nigeria	2008
4	1	18	Compacta	2009
5	3	6	Compacta	2009
6	3	19	Deli X Ghana	2010
7	1	12	Deli X Ghana	2009
		13	Deli X Nigeria	2009
8	1	2	Deli X La mé	2008
9	5	22	Deli X La mé	2011
10	5	38	Deli X Nigeria	2012
11	2	9	Deli X Ghana	2008
12	1	2	Deli X La mé	2008
13	1	4	Deli X La mé	2008
		5	Deli X La mé	2008
14	4	33	Deli X Ghana	2011
15	4	39	Deli X Nigeria	2012

Fuente: elaboración propia.

Tal como se puede apreciar en la tabla anterior, los racimos de un centro de acopio no necesariamente contienen un mismo lote de procedencia, ya que como se describe para el pesaje 1, 2, 7 y 13 el centro de acopio recolectó fruta de dos lotes diferentes. Es importante aclarar que en el formato de registro se hizo diferenciación de los racimos procedentes de un lote y de otro.

- Comparación de métodos de muestreo en tolva de planta de beneficio

A 5 de las 15 poblaciones de racimos pesados en el centro de acopio se les dio seguimiento hasta su llegada a la rampa de recepción de fruta en planta de beneficio, lugar en el que se realizó un muestreo de 30 racimos, por medio de dos metodologías de muestreo.

Las unidades de transporte depositaron los racimos y la toma de peso de las muestras fue realizada de la forma ilustrada en la figura 2.

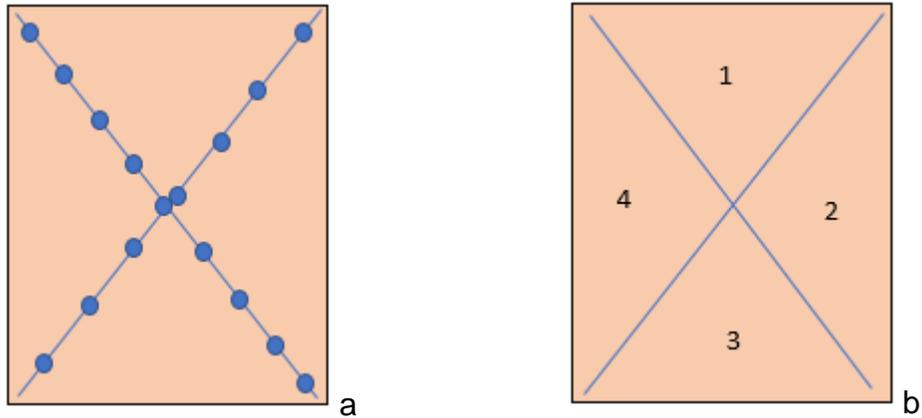
Figura 2. **Racimos en planta de beneficio a) rampa; b) peso de la muestra**



Fuente: elaboración propia.

Tal como se mencionó anteriormente se compararon dos métodos para selección de muestra. El primer método utilizado es conocido como el método de la cuerda, el cual consiste en una cuerda de 6 m con nudos equidistantes a cada 40 cm, debido a que la cuerda posee 15 nudos se lanza 2 veces en una secuencia que tenga forma de X, los racimos que coinciden con los nudos son tomados como parte de la muestra hasta completar un total de 30 racimos. El segundo método evaluado consistió en seccionar el contenido de la unidad de transporte en 4 cuadrantes, dónde por medio de una la misma cuerda con nudos se seleccionaron 8 racimos en el primer y tercer cuadrante y 7 racimos en el segundo y cuarto. Ambas metodologías fueron realizadas según la manera en que se representa en el esquema de la figura 3.

Figura 3. **Método de muestreo a) cuerda; b) cuadrante**



Fuente: elaboración propia.

Para comparar la representatividad de cada método se realizaron 5 muestreos de calidad de fruta en tolva de planta de beneficio.

- Fase III. Análisis de la información

El registro y almacenamiento digital de la información se realizó a través del programa Excel. El análisis estadístico de las pruebas de bondad de ajuste para comprobar la normalidad del peso de racimos y el contraste de medias fue realizado por medio de los programas estadísticos Rstudio e Infostat.

- Determinación del tamaño muestral requerido para un muestreo representativo y significativo de la calidad de racimos

Inicialmente se comprobó que el peso de racimos se ajustara a una distribución de probabilidad normal por medio de la prueba de Kolmogorov-Smirnov (K-S), con un nivel de significancia del 5 %. Una vez comprobada la normalidad del peso de racimos se procedió al cálculo de media y varianza, en cada pesaje para poder emplear la fórmula de muestreo irrestricto aleatorio, descrita por Scheaffer y Mendenhall en 1997 en la que se utilizó un nivel de imprecisión de 3 kg.

- Contraste de hipótesis para dos métodos de muestreo en tolva de planta de beneficio

Para conocer la representatividad de cada método de muestreo se comparó la media y varianza poblacional con la media y varianza muestral de ambos métodos, también se realizó una prueba de comparación de medias t Student, en la que se determinó la diferencia estadística de las metodologías evaluadas.

INTRODUCCIÓN

Los racimos de palma de aceite son la materia prima para gran variedad de productos comercializados en las industrias de biocombustible, cosmética femenina y producción de alimentos, razón por la cual las empresas dedicadas a la palmicultura se someten a procesos de certificación que aseguren la calidad de la producción.

Como parte de los procesos internos de calidad, se realizan muestreos de racimos provenientes de campo, previo al ingreso de la planta de beneficio y se identifica el grado de maduración de este, ya que ese es el principal factor del cual depende la calidad del aceite producido.

Debido a la carencia de investigaciones que se desarrollen entorno a la medición de calidad de racimos específica para Guatemala y sus regiones productivas, se han tomado como referencia estudios realizados en Colombia. Como principal aporte de la presente investigación se validó, desde el enfoque de estudio de muestreo estadístico, el proceso de selección y el número de muestra utilizados para las mediciones de calidad con información generada en la región de la Franja Transversal del Norte de Guatemala.

El presente documento consta de 4 capítulos, el primero describe los antecedentes y el marco conceptual bajo el cual se desarrolló el estudio; en el segundo se describen todos los fundamentos agrícolas y estadísticos que se relacionan con el tema de estudio y que son la base para la explicación de los resultados; el tercero detalla los resultados obtenidos y en el cuarto se discute acerca de los estadísticos obtenidos y los factores influyentes para los mismos.

Como resultado del análisis de información contenida en el primer capítulo se estableció la metodología del estudio. Se realizaron mediciones en campo dirigidas a unidades de transporte y se determinó el número de muestra requerido para un nivel de significancia del 5 %, se calcularon los intervalos de límite para el error y se realizaron algunas correlaciones que permitieron validar que el número mínimo de muestra es de 27 racimos. Así mismo se realizaron evaluaciones para comparar dos métodos de selección de racimos en donde se determinó que no existen diferencias estadísticas entre el método de la cuerda y el método del cuadrante.

Los resultados del estudio realizado demuestran que el peso de las poblaciones de racimos no siempre se ajusta a la distribución normal, ya que algunas distribuciones presentan colas muy largas y es necesario aplicar los conceptos de transformación exponencial, sin que estos procedimientos aseguren cumplir el supuesto de normalidad.

Las poblaciones que presentaron poca congruencia con la distribución normal fueron las que tuvieron mayor cantidad de fruto suelto, lo que sugiere mantener buenas prácticas de cosecha, alza y transporte para mantener la normalidad del peso de racimos. Sin tomar en cuenta los casos en los que no se logró comprobar la normalidad del peso de racimos los resultados sugieren que al menos para la región de la Franja Transversal del Norte, es válido tomar una muestra de 30 racimos y el método de la cuerda como forma de selección de racimos.

1. MARCO REFERENCIAL

1.1. Estudios Previos

Todas las operaciones agrícolas desarrolladas en las plantaciones de cultivos como la palma de aceite tienen como objetivo cumplir metas de productividad y garantizar la calidad de la producción, debido a que de ello dependen la rentabilidad del negocio y las características del aceite producido en el sitio de beneficio.

Los procesos de certificación y mejora continua han generado la necesidad de medir la de calidad de fruta fresca como parte de la garantía en la calidad del aceite desde la materia prima. Entorno a ello, países como Malasia y Colombia, en donde existen estudios con dicho enfoque, han generado información que puede ser utilizada como guía para estudios en Guatemala.

Las mediciones de calidad de la fruta se realizan en la planta de beneficio y se enfocan en las unidades de transporte que trasladan la fruta de las plantaciones al proceso de beneficio, por lo que los muestreos son dirigidos a dichas unidades de transporte.

La calidad de los racimos podría ser evaluada en campo, pero esta evaluación tiene al menos 3 limitantes:

Requiere de gran cantidad de recursos debido a que una plantación puede tener múltiples frentes de cosecha. La evaluación no representa la calidad de la fruta al ingresar al proceso ya que podría ser golpeada o demorada

en el transporte, afectando su calidad y cuando la fruta es provista por un tercero a una planta de beneficio, la evaluación en el campo no es viable por la distancia y dispersión de las plantaciones a visitar. Por tanto, el proceso de evaluación deberá llevarse a cabo en la planta de beneficio. (Erales, 2016, p. 11).

En la metodología del presente estudio se tomaron en consideración los aspectos mencionados por Erales (2016) ya que sus argumentos justifican que la evaluación de calidad de fruta sea realizada en la rampa de recepción de la planta de beneficio, previo al proceso de esta y no en campo.

La metodología para determinación del número de muestra se basó principalmente en los estudios realizados por García y Yáñez (2000) quienes determinaron el número mínimo de racimos requerido para el muestreo bajo las condiciones del Norte y Occidente de la zona palmera de Colombia. De este estudio se adoptó el parámetro de muestreo, el nivel de significancia y error de muestreo por ser congruentes con la media y la varianza de las poblaciones evaluadas en el estudio realizado en Fray Bartolomé de las Casas.

El método de muestreo empleado en la región de estudio está basado en la investigación de García y Yáñez (2000), razón por la cual una de las dos metodologías de selección de muestra evaluadas en Fray Bartolomé de las Casas (la cuerda), también se encuentra descrita en el mismo estudio.

Previo al año 2000, existen otras publicaciones que hacen referencia al número de muestra requerido para las evaluaciones de calidad, por ejemplo, Toong y Yang (1993) especifican que es necesario seleccionar por lo menos 100 racimos recién cortados. García y Yáñez (2000) citan a Porla y Abdul quienes mencionan que se requieren aproximadamente 10 racimos/tonelada

seleccionados aleatoriamente. En torno a este sentido, *The Thai Agricultural Standard* (2008), refiere que cuando la unidad de transporte sea menor a 50 Tm, la muestra debe ser de 50 racimos y en caso de que la unidad de transporte sea igual o mayor a 50 Tm deben seleccionarse 100 racimos, ya que “un volumen alto de racimos elimina la necesidad de establecer un método de muestreo” (García y Yáñez, 2000, p.307).

Por la carencia de sustento estadístico que se expone los estudios de Toong y Yang, *The Thai Agricultural Standard* (TAS), Porla y Abdul no fueron considerados como base en la formulación de la metodología, sino que se tomaron como referencia para la discusión y comparación del número de muestra propuesto en documentos previos y los resultados del presente.

Respecto a la metodología para la toma de la muestra se menciona en los mismos trabajos citados anteriormente, que los racimos deben ser extraídos de forma aleatoria pero no hacen referencia a la forma de selección. En este caso, Erales (2016) y García y Yáñez (2000) refieren la metodología de la cuerda con nudos equidistantes a cada 0.7 m o 0.4 m, en la que se explica que la cuerda debe ser lanzada dos veces a manera de formar una X y abarcar la totalidad de la tolva, (debe procurarse que el largo de la cuerda permita la selección de 15 racimos en cada lanzamiento).

García y Yáñez (2000), también mencionan la metodología de la vara y la canasta, la primera consiste en una vara de 1 m para la selección de los racimos con uno de sus extremos mientras se realiza un recorrido en forma de X. La segunda técnica de selección de los racimos es propuesta durante el llenado de la unidad de esterilización. Para los propósitos planteados en este documento no se consideraron las últimas dos metodologías por ser poco prácticas y faltar a algunos preceptos de muestreo probabilístico.

Para conocer el número de racimos requerido y conformar una muestra en la evaluación de la calidad de fruta de palma de aceite en Guatemala, se validó la metodología de la cuerda propuesta por Erales (2016) y García y Yáñez (2000) y se comparó con la metodología del cuadrante, que fue propuesta como resultado de las recomendaciones del personal de evaluación de Fray Bartolomé de las Casas y es considerada una variante de la metodología de la cuerda.

Para el análisis de normalidad se tomó como referencia a Bracamonte (2018), Romero (2016), Jardón (2007), Márquez (2010) y Vela (2010) quienes explican el ajuste de una variable aleatoria a la distribución normal por medio de pruebas como Kolmogorov-Smirnov (K-S), Test de Lilliefors, Jarque-Bera y pruebas gráficas. Conceptos que fueron utilizados para la interpretación de los resultados en la distribución de probabilidad del peso de racimos.

Documentos como los de Rodríguez, Ferreras y Núñez (1991), también fueron utilizados para el sustento estadístico del presente documento, ya que refieren la importancia de conocer los niveles de precisión en un diseño muestral para la inferencia estadística. De tal forma que sugieren el uso de indicadores como el coeficiente de variación y el error de muestreo para sustentar la certeza y calidad del método evaluado. En este sentido Vecco, Pinedo y Fernández (2015) en el artículo que aborda el muestreo de poblaciones de *trips* (Insectos del orden Thysanoptera, plaga de varios cultivos de interés económico para el ser humano) y sus enemigos naturales toman en cuenta dicha precisión estadística al realizar regresiones lineales para los estimadores y proponer el uso de márgenes de error para el muestreo.

1.2. Marco contextual

Las fincas y lotes tomados en cuenta para el estudio son jurisdicción del municipio de Fray Bartolomé de las Casas, Alta Verapaz, Guatemala, cuya cabecera se encuentra ubicada en las coordenadas UTM 191986.97 Latitud Norte y 1749824.95 Longitud Oeste (Sinsinaj, 2007).

- Condiciones Climáticas

El municipio de Fray Bartolomé de las Casas posee un clima cálido, en donde Sinsinaj (2007) indica que la temperatura promedio es de 26 °C, con temperaturas máximas de 38 °C y registros de temperaturas mínimas de 14 °C.

La humedad relativa promedio es de 82 %, la época lluviosa no se encuentra bien definida, pero la precipitación dura alrededor de 8 meses, con un promedio anual de 2332 mm (Sinsinaj 2007).

- Zona de vida

Según Sandoval (2007) de acuerdo con el sistema de clasificación de Holdridge, el área de estudio se encuentra clasificada dentro de la zona de vida de Bosque muy Húmedo subtropical cálido. Representado por las siglas bmh- S (c).

- Historia de la palma de aceite en Guatemala

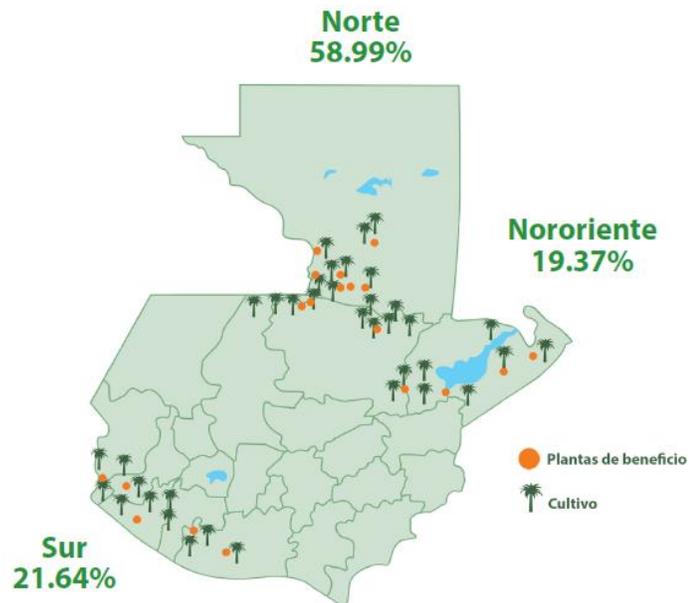
En la década de los 40 la *United Fruit Company* incursionó en el establecimiento de las primeras plantaciones de palma de aceite en Guatemala.

Luego de unos años de cultivo, dichas plantaciones fueron abandonadas por temas de reforma agraria (Grepalma, 2017).

Posteriormente alrededor del año 1960, la caficultura en Guatemala atravesó por la crisis que propició que la Asociación Nacional del Café (ANACAFE) estableciera 225 hectáreas de palma de aceite y una planta de beneficio en el sur del país (Grepalma, 2017). Sin embargo, fue hasta en los años de 1980 durante la crisis del algodón, cuando se establecieron las primeras plantaciones comerciales en los departamentos de Escuintla y San Marcos (Grepalma, 2017).

Información publicada por el CABI (2016) y Grepalma (2017) especifica que los departamentos influenciados por el cultivo e industria de la palma aceitera son: San Marcos, Quetzaltenango, Retalhuleu, Suchitepéquez, Escuintla, Izabal y Alta Verapaz (en los valles del río Motagua y Polochic) y Quiché. Estos departamentos se agrupan en tres regiones que representan 165,511 hectáreas de cultivo de palma en Guatemala. La región de la costa sur representa 21.64 % del área cultivada, la región nororiente al 19.37 % y finalmente la región norte, que es la más grande a 58.99 % del total de área cultivada (figura 5).

Figura 4. **Distribución del cultivo y plantas de beneficio en Guatemala**



Fuente: GREPALMA (2017). *I Anuario Estadístico 2016-2017* (p. 3).

Para el año 2016 Guatemala se situó como el tercer país con mayor exportación de aceite crudo a nivel mundial, por debajo de Indonesia y Malasia. El principal mercado hacia el que se dirige el aceite guatemalteco es México, Países Bajos y Alemania (CABI, 2016).

También se debe resaltar que la actividad palmera tiene un impacto social y económico en el país, ya que según GREPALMA (2017) la actividad genera hasta 28,000 empleos directos y representa el 1.2 % del Producto Interno Bruto (PIB).

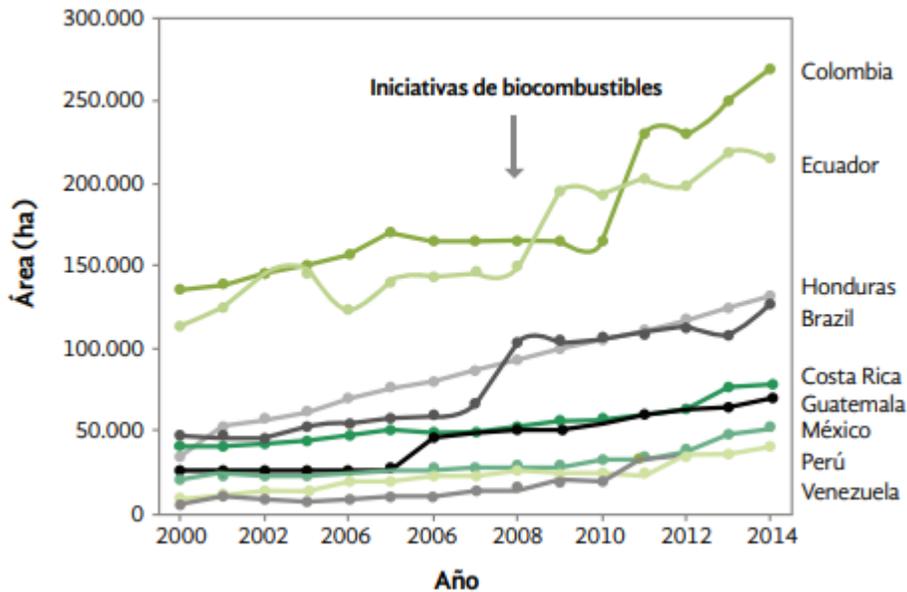
2. MARCO TEÓRICO

2.1. Palma de aceite

Según registros históricos la palma de aceite fue mencionada por primera vez por Eannes, quien fue comisionado por Henry, *El navegante*, para rodear el cabo de Bojador y explorar la Costa Occidental de África. El nombre científico aceptado y por medio del cual se le reconoce a la palma es *Elaeis guineensis* Jaqc y fue nombrada así por Jacquin en 1763. El nombre científico de la palma africana deriva de la palabra *Elaeis*, del griego *elaion*, que significa aceite y el epíteto específico *guineensis*, denota que Jacquin atribuía su origen a la costa de guinea (Corley y Tinker, 2003 y Fairhust y Härdter, 2012).

Tal como se ha descrito anteriormente, la palma de aceite se originó en África, pero es en Asia en dónde se encuentran las mayores extensiones de este cultivo, especialmente en Indonesia y Malasia. Latinoamérica también representa un sector productivo para la palma, según la FAO los países con mayor extensión de palma de aceite son Colombia y Ecuador, la curva de crecimiento en la figura 5 muestra incremento de área de cultivo a partir del año 2008 debido, principalmente a la influencia del biocombustible (Forumo y Mitchell, 2017).

Figura 5. **Área cultivada en América Latina (2000 -2014)**



Fuente: Forumo y Mitchell (2017). *Caracterización de la expansión de la palma de aceite para uso comercial en América Latina: cambio en el uso del suelo y comercialización* (p. 30).

La producción de aceite de palma incrementó, en promedio, 58 % en un período de 7 años comprendido de 1994 al 2001, situación que atribuyeron a su vez, al incremento de la demanda como consecuencia del crecimiento poblacional y económico mundial (Fairhurst y Härdter, 2012). Aunado a la creciente demanda algunos expertos como Forumo y Mitchell (2017) también reportan incremento en la producción de aceite y explican que se debe al crecimiento de área cultivada y a la mejora de la productividad de las plantaciones.

La palma de aceite es conocida por su capacidad no superada para captar y transformar la energía solar en aceite vegetal (Fairhurst y Härdter, 2012). La palma de aceite aporta el 33 % de grasas vegetales en el mercado mundial, pero

ocupa para ello el 8 % del área cultivada, por lo tanto, lo que representa un rendimiento de 0.3 ha/TM aceite.

2.1.1. Taxonómica y morfología

The Plant List (2018) y Corley y Tinker (2003) citan la clasificación taxonómica de *Cronquist* que describe a la palma de aceite como una planta que pertenece a la familia *Arecaceae* (conocido anteriormente como *Palmae*) y al orden *Arecales*.

Elaeis guineensis Jacq es una palma grande que se caracteriza por poseer un solo tallo columnar y entrenudos cortos. Sus hojas son pinnadas simples, pueden medir entre 7 y 8 m de largo con filotaxis en espiral. Es una especie monoica, sus flores se desarrollan en las axilas de las hojas y los frutos son un conjunto de frutos que se encuentran adheridos a un mismo tallo a los que se les denomina racimos (Corley y Tinker, 2003 y Fairhurst y Härdter, 2012).

Para la industria de cosméticos femeninos, biocombustibles y alimentos el órgano de interés económico es el fruto, éste es conocido comúnmente como racimo y es una drupa sésil de forma casi ovoide (Corley y Tinker, 2003). Un racimo puede contener alrededor de 1500 frutos, cuando está maduro su peso puede variar entre 3 -10 kg en palmas jóvenes y 10-30 kg en palmas adultas. Los frutos de donde se extrae el aceite están formados por un exocarpio, un mesocarpio aceitoso, un endocarpio duro y un endospermo llamado almendra. El aceite se obtiene del mesocarpio y contiene alrededor del 11-21 % del material fibroso (Fairhurst y Härdter, 2012 y Corley y Tinker, 2003).

Se puede encontrar aceite en los frutos exteriores, que poseen mayor pigmento y también en los frutos interiores que están más comprimidos. También

existen frutos partenocárpicos que se han desarrollado sin fertilización, pero no producen aceite (Fairhurst y Härdter, 2012).

2.2. Medición de calidad de racimos

La medición de calidad de fruta es necesaria, debido a que las plantas de beneficio necesitan extraer aceite de alta calidad. Las características requeridas por las plantas refinadoras de aceite son: acidez, humedad, blanqueabilidad, oxidación, contenido de monodiglicéridos, contenido de diglicéridos y contenido de antioxidantes naturales (Taillez, Siaka, Bonny y Jacquemard, 1996 y Mohd 2002). Estas características están relacionadas directamente con la maduración de los racimos, la cantidad de frutos sueltos y la coloración de los racimos pueden ser utilizadas como indicadores de maduración.

Taillez et al., (1996) definen que el contenido de aceite que posee la pulpa de la fruta desprendida es más bajo que el de los frutos que permanecen adheridos al racimo. Por tal razón, se recomienda que el corte de racimos sea tan pronto como se desprendan los primeros frutos.

En torno a este tema, las empresas dedicadas a la palmicultura en Guatemala poseen criterios de cosecha en los que se recomienda organizar la actividad de corte en periodos de aproximadamente 15 días, ya que es el tiempo alrededor del cual se desprenden de 3 a 5 frutos. El número de frutos desprendidos es un criterio que las empresas escogen según sus intereses y que puede verse afectado por la fisiología de la palma y como consecuencia de condiciones climática. Esta práctica como ya se mencionó favorece la calidad de aceite, permite reducir las pérdidas y costos de cosecha, minimiza el contenido de impurezas y también reduce la acidez del aceite (Erales, 2016; Taillez et al., 1996).

La apariencia externa del fruto varía, según Fairhurst y Härdter (2012) se pueden describir principalmente dos tipos de fruto, el tipo más común que es de color violeta a negro en el ápice y descolorido en la base, previo a la madurez y que una vez en estado maduro el color varía de anaranjado a rojo, probablemente debido a los cambios en el contenido de carotenos. El otro tipo de fruto es el que es verde previo a la madurez y luego cambia a un color ligeramente anaranjado rojizo.

Para la calificación de racimos de palma de aceite, Malasia ha generado estándares de clasificación de madurez como la descrita por *The Malaysian Palm Oil Board* (MPOB) (2006) y Abdul, Zulkifli, Rahman y Salleh (2009), dichos estándares se encuentran basados principalmente en la coloración de los racimos y en el porcentaje de desprendimiento de fruto, los cuales están relacionados directamente con la madurez de estos.

Con base en la síntesis realizada por Hazir, Shariff y Amiruddin (2012) en dónde se elaboró la tabla 3 que es un resumen de las principales características y categorías detalladas por MPOB (2006) y Abdul et al. (2009), la tabla describe el estado de madurez de los racimos de palma de acuerdo con criterios de coloración del mesocarpio y alveolos vacíos.

Tabla III. **Clasificación de racimos basada en frutos sueltos y color**

Número de alvéolos vacíos	Color del mesocarpio		
	Amarillo	Amarillo-naranja	Naranja
0	Inmaduro	Inmaduro	Maduro
0-10	Inmaduro	Poco-maduro	Maduro
>10	Inmaduro	Maduro	Maduro

Fuente: Hazir M., Shariff A. y Amiruddin M. (2012). *Determination of oil palm fresh fruit bunch ripness-Based on flavonoids and anthocyanin content* (p. 486).

Hazir et al. (2012) plantean otro estándar de clasificación de acuerdo con el número de alvéolos vacíos, los cuales pueden ser utilizados para incrementar la exactitud de la clasificación, tal como se describe en la tabla 4.

Tabla IV. **Clasificación de madurez de racimos con base en fruto suelto**

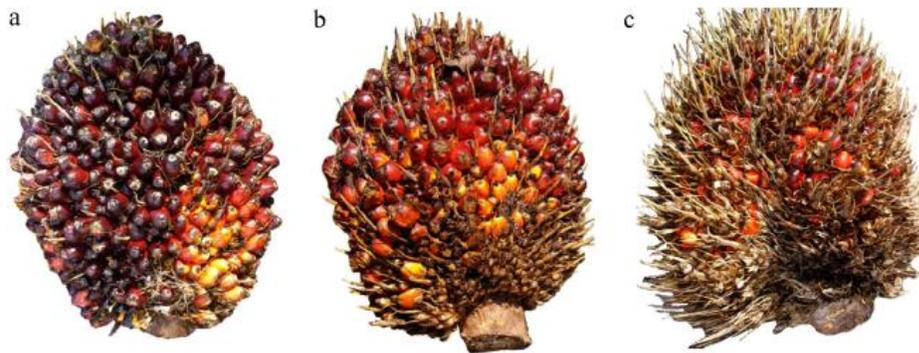
Categoría	Descripción
Maduro	Fruto suelto entre 10 % y 50 %
Sobre maduro	Fruto suelto entre 50 % y 90 %
Inmaduro	De 1 a 9 frutas sueltas

Fuente: *Determination of oil palm fresh fruit bunch ripness-Based on flavonoids and anthocyanin content*. Adaptado de Hazir M., Shariff A y Amiruddin M. (2012, p.486).

En Guatemala la evaluación de calidad de racimos se realiza de forma visual, debido a que existen casos en los que no es sencillo reconocer la madurez de un racimo, se recomienda utilizar el sistema de clasificación descrito en la tabla 3 y reforzarlo con la ilustración de la figura 6, ya que al combinar dos características de maduración (color y número de alveolos vacíos) se fortalece el

criterio en comparación a la descripción de la tabla 4, donde solo se utiliza el porcentaje de alveolos vacíos.

Figura 6. **Categorización de fruta; a) verde; b) maduro; c) sobre maduro**



Fuente: Hazir M., Shariff A. y Amiruddin M. (2012). *Determination of oil palm fresh fruit bunch ripeness-Based on flavonoids and anthocyanin content* (p. 486).

Las clasificaciones anteriormente mencionadas son las más comunes, pero existen estudios contemporáneos que mediante la incorporación de tecnología permiten clasificar la madurez de los racimos a través de métodos más objetivos, entre los que se pueden mencionar: el uso de fluorescencia para detección de flavonoides y antocianinas, la perforación y medición de fuerza de inyección en el mesocarpio de los frutos, la utilización de software para análisis de colores rojo, verde y azul, el escaneo de fruta por medio de tecnología hiperespectral y de infrarrojo cercano y espectroscopia de fluorescencia e imágenes térmicas (Hazir et al. , 2012; Rashid, 2016; Azli, Shahril, Nor Jayati y Sukanur, 2009; Meftha, Abdul, Mohd, Saared y Eshanta, 2008).

Un racimo que no cumple con los estándares de madurez es penalizado en las evaluaciones de calidad, debido a que el verde no ha llegado a su máxima

capacidad de síntesis de aceite y el sobre maduro compromete la calidad de este. Por sus características favorables para el negocio de palma, las evaluaciones de calidad de racimos buscan que la totalidad de éstos concuerden con la descripción de racimo maduro, tal como se describe en la tabla 4 (Clasificación de madurez de racimos con base en fruto suelto) y la figura 6 (Categorización de fruta fresca).

Para ampliar el tema de los racimos sobre maduros, estos tienden a acumular de ácidos grasos libres, que tienen influencia en la cadena de eventos que contaminan el aceite. Además, también afectan los procesos de oxidación y algunas condiciones físicas del aceite (Fabio y Calvo, 1991).

2.3. Fundamentos Estadísticos

La palabra estadística deriva del vocablo *Statistik* que tiene origen en la palabra italiana *statista*, la primera persona en utilizar el término fue Gottfried Achenwall, pero se popularizó a través del informe estadístico sobre Escocia (1791-1799), en el informe se define a la estadística como la ciencia que proporciona métodos, técnicas y procedimientos para el análisis de datos con el fin de describirlos y/o emitir generalizaciones válidas (Levin & Rubin, 2004).

Levin y Rubin (2004) clasifican la estadística en descriptiva e inferencial, el primer término hace referencia a los métodos para describir poblaciones y el segundo a los que se emplean con el fin de realizar inferencias.

2.4. Estadística Inferencial

A través de los métodos que conforman este tipo de estadística, es posible obtener información de una población sin necesidad de analizar todos los

individuos que la conforman. Córdova (2003) indica que para lograrlo se deben aplicar los principios de azar a la selección de individuos que conformarán la muestra y que las mediciones realizadas a dicha muestra se relacionan con la población para poderla describir.

Levin y Rubin (2004) agregan que para brindar generalizaciones acertadas es necesario obtener una muestra válida, objetivo que se logra a través de los conocimientos de la teoría de probabilidades y el empleo de parámetros que midan la confiabilidad del muestreo.

2.4.1. Probabilidad

El concepto de probabilidad matemática puede ser descrito como la idea de seguridad de aparición de un suceso aleatorio (Maibaum, 1988). Esta idea de aparición de sucesos se define por el cociente:

$$P(A) = \frac{g(A)}{K} \quad (\text{Ec. 1})$$

dónde:

$P(A)$ = probabilidad de aparición de un suceso aleatorio

$g(A)$ = número de resultados favorables para un suceso aleatorio

K = Número total de resultados

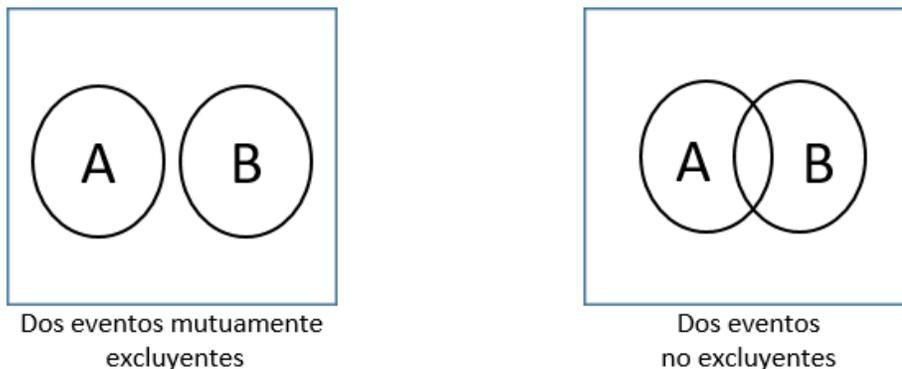
Maibaum (1988) señala que la probabilidad de un suceso puede ser expresada de forma absoluta o relativa en una distribución de frecuencias, para los fines de la inferencia se recomienda hacerlo de forma relativa y se expresa por medio de porcentajes.

2.4.2. Espacio muestral

Al conjunto de posibles resultados en un determinado proceso se le conoce como espacio muestral. Los estadísticos hacen uso de la palabra experimento para referirse a un proceso que genere datos, por lo tanto, para un experimento el espacio muestral sería llamado espacio muestral experimental (Levin y Rubin, 2004 y Walpole et al., 2012). Para ilustración de las ideas planteadas se denota como $S: \{H, T\}$ donde S es el espacio muestral y H y T son eventos cualesquiera.

El espacio muestral (S) puede ser representado mediante diagramas de Venn, en donde para el caso representado en la figura 7 el espacio muestral completo es delimitado con un rectángulo y un evento se representa como parte del rectángulo. Para eventos mutuamente excluyentes y no excluyentes se representa de la siguiente manera:

Figura 7. **Espacio muestral de eventos excluyentes y no excluyentes**



Fuente: Levin y Rubin (2004). *Regla de la adición para eventos mutuamente excluyentes* (p. 138).

Para eventos excluyentes mutuamente se expresa como:

$$P(A \text{ o } B) = \text{Probabilidad de que suceda } A \text{ o } B$$

y se calcula:

$$P(A \text{ o } B) = P(A) + P(B) \quad (\text{Ec. 2})$$

Levin y Rubin (2004) expresan que esta suma de probabilidades es conocida como regla de adición para eventos que se excluyen mutuamente y que para el caso de regla de adición para eventos que no son mutuamente excluyentes, la regla adquiere la siguiente modificación debido a la posibilidad de que ambos eventos se representen simultáneamente:

$$P(A \text{ o } B) = P(A) + P(B) - P(AB) \quad (\text{Ec. 3})$$

dónde:

$P(AB)$: Probabilidad de que A y B sucedan a la vez.

2.5. Muestreo estadístico

La estadística pretende describir poblaciones a partir de su distribución de probabilidad. En general, factores de índole logístico, temporal y/o económico complican el censo y hacen que sea necesario utilizar el muestreo estadístico. El muestreo estadístico permite encontrar el número de individuos mínimo necesario para realizar inferencias significativas y representativas a través de estimadores como la media, mediana, varianza, desviación estándar y otros. (Scheaffer y Mendenhall ,1997 y Días, Ojeda y Valderrábano, 2016).

En el entorno del muestreo probabilístico Días, Ojeda y Valderrábano (2016) describen como población objetivo o de referencia al conjunto de individuos del cual interesa conocer alguna de sus características, dicho conjunto debe delimitarse conforme los intereses del investigador y se puede representar como $U = \{x/x \text{ es un elemento de la población}\}$, en dónde, para el caso de las poblaciones infinitas se desconoce cuántos individuos conforman la población.

Las inferencias realizadas por la estadística dirigidas a la estimación de características numéricas de una población, tales como media, total o varianza, son conocidas como parámetros (Días et al., 2016).

Scheaffer y Mendenhall (1997) enfatizan sobre la importancia de modelar la distribución de probabilidades teórica asociada a los datos observados previo al análisis inferencial de la información, ya que la probabilidad reflejada en dichos modelos es indispensable para la selección de procedimientos estadísticos válidos.

Debido a que el objetivo del muestreo es obtener un subconjunto representativo Lohr (1999) también sugiere considerar los siguientes aspectos para eliminar el sesgo:

- No utilizar criterios de selección que estén asociados a la característica de interés,
- A no ser que sea parte del propósito del estudio, no seleccionar elementos de forma intencional,
- Evitar la ambigüedad en la delimitación de la población objetivo,
- Cuidar la subcobertura de la población objetivo,
- No sustituir elementos de muestreo por otros más convenientes,

- Evaluar los campos vacíos, previo al análisis matemático de la información.

Las recomendaciones anteriormente listadas son válidas a menos que el muestreo sea del tipo no probabilístico (Días et al., 2016).

Se les conoce como parámetros poblacionales a las medidas de tendencia central y dispersión que resultan de un censo poblacional, dentro estas medidas las más importantes son la media y la varianza, mismas que cuando son calculadas con base en la inferencia se les denomina estimadores (Scheaffer y Mendenhall, 1997).

Días et al. (2016) indican que los mecanismos de azar son fundamentales en la construcción de frecuencias, se debe evitar el sesgo y reducir el error de muestreo para que los valores se aproximen a su valor teórico esperado y a los criterios de probabilidad normal.

2.5.1. Tipos de muestreo

Para obtener una muestra representativa y significativa existen varios métodos, estos deben ser seleccionadas por el investigador conforme a las necesidades planteadas por el estudio y a las circunstancias de muestreo bajo las que se encuentra sujeto el investigador. Algunos de los métodos utilizados para selección del tamaño de una muestra son: el muestreo aleatorio estratificado, estimación de razón, regresión y diferencia, muestreo sistemático, muestreo por conglomerados, muestreo por conglomerados en dos etapas y el muestreo simple aleatorio (Sheaffer y Mendenhall, 1997).

Dentro de los cuales “los tipos de muestreo básicos son: muestra simple aleatoria, muestra estratificada y muestra por conglomerados” (Lohr, 1999, p.24).

- Muestreo simple aleatorio

Se considera que este tipo de muestreo es la base del muestreo probabilístico ya que el resto de las metodologías se derivan o tienen su origen a partir de este. El objetivo es seleccionar un subgrupo n a partir del grupo total N , nunca se debe olvidar que el aspecto más importante para la efectividad de este y cualquier otro tipo de muestreo es reducir el sesgo en su mínima expresión a través del respeto de los mecanismos de azar (Sheaffer y Mendenhall, 1997).

La fórmula utilizada y descrita por Sheaffer y Mendenhall (1997) para determinar el tamaño de la muestra para un muestreo simple aleatorio es:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad (\text{Ec. 4})$$

dónde:

n = número de muestra

N = tamaño de la población

σ^2 = varianza

$D = \frac{B^2}{4}$ y B corresponde al límite para el error de estimación

2.6. Distribución de probabilidad

Una distribución de probabilidad es un cálculo estimado del comportamiento de un fenómeno que se modela a través de frecuencias relativas, cada fenómeno observado en la naturaleza posee características que lo hacen similar a una distribución teórica, en el caso en que se comprueben estadísticamente dichas similitudes se pueden utilizar los conceptos ya estudiados de las distribuciones teóricas para explicar los sucesos observados (Lohr, 1999).

Levin y Rubin (2004) explican que todos los fenómenos motivo de estudio para el ser humano ofrecen cierto grado de incertidumbre, aun cuando se aplican los aportes de las ciencias exactas, para minimizar esa incertidumbre y no caminar a ciegas al momento de tomar decisiones se debe buscar la distribución teórica que mejor ajuste tenga con los datos del evento que esté en estudio.

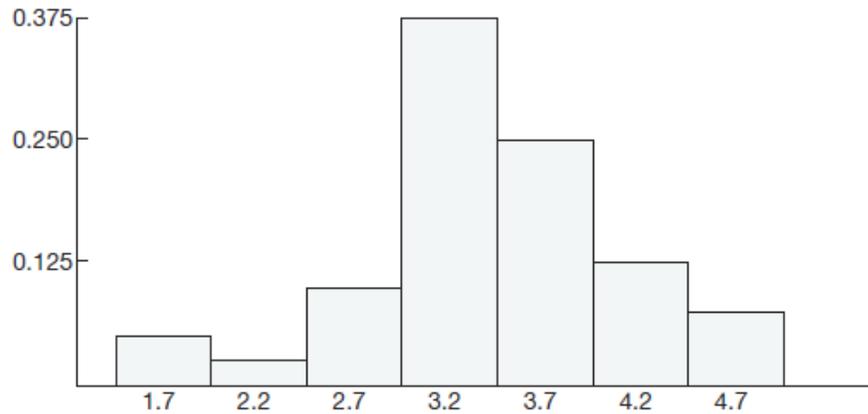
2.6.1. Distribución de frecuencias

Este es un método de representación que se utiliza para variables continuas, aunque para ciertas circunstancias también puede ser utilizada para variables discretas (Córdova, 2003). Una distribución de frecuencias ordena valores para indicar que tan a menudo aparecen en un conjunto de datos. La distribución de frecuencias se puede calcular de forma absoluta, relativa y de porcentaje (Córdova, 2003; Levin y Rubin, 2004).

2.7. Histograma

Es una representación gráfica de una distribución de frecuencias ya sean relativas, absolutas o de porcentaje. La ilustración consiste en barras verticales y juntas en donde la base de cada barra es proporcional al intervalo y la altura a la frecuencia de los valores. La figura 8 muestra la forma de un histograma, en donde el eje horizontal se representa la escala de la variable y el eje vertical las frecuencias (Córdova, 2003).

Figura 8. **Histograma**



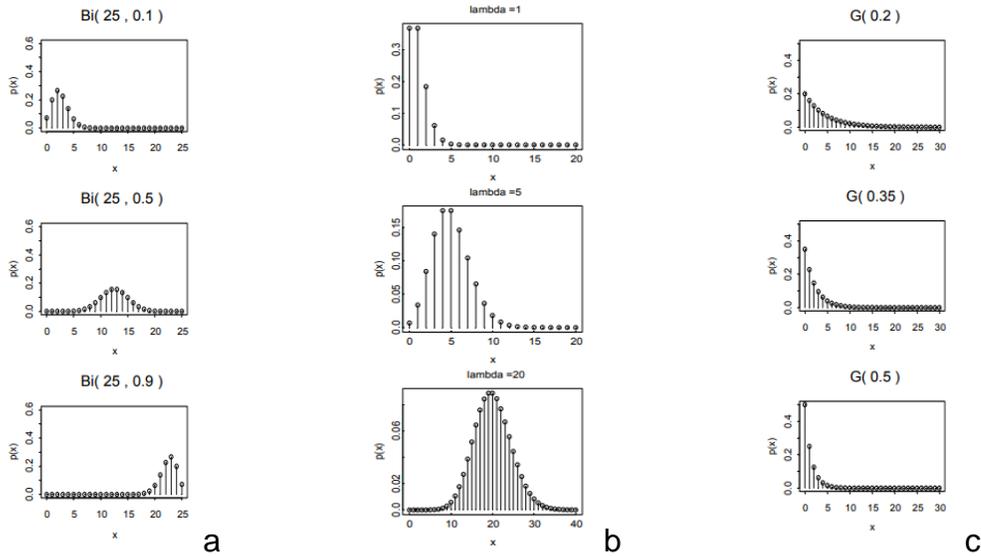
Fuente: Walpole, Myers, Myers y Ye (2012). *Histograma* (p. 23).

El histograma es un método que puede ser utilizado para la estimación empírica de la forma probabilística de una población y para analizar otras características como simetría y sesgo (Scheaffer y Mendenhall, 1997; Walpole et al., 2012).

2.8. Distribuciones para variables discretas

Estas distribuciones son utilizadas para describir variables aleatorias. En este tipo de distribución se usa un número limitado de valores y asigna una probabilidad de ocurrencia entre el límite superior y el inferior de la variable, Dentro de las distribuciones para variables discretas se pueden mencionar la distribución binomial, Poisson e hipergeométrica, principalmente (Walpole et al., 2012).

Figura 9. Distribución a) Binomial; b) Poisson; c) Hipergeométrica

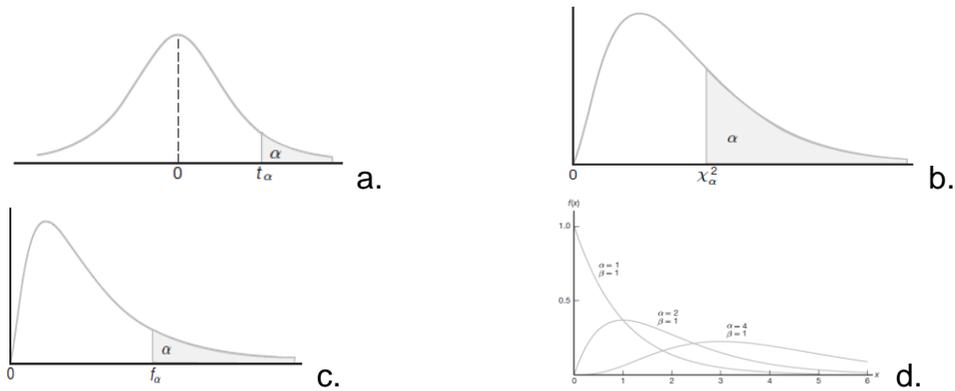


Fuente: Bianco y Martínez (2004). *Variables aleatorias discretas* (p. 41, 43 y 50).

2.9. Distribución para variables continuas

La variable considerada toma cualquier valor dentro de un intervalo. Dentro de las distribuciones continuas se pueden nombrar la distribución uniforme, gamma, exponencial, chi cuadrada, t – Student, F-Snedecor, beta, logarítmica normal, weibiull y la distribución normal (figura 10).

Figura 10. **Distribución a) t Student; b) χ^2 ; c) F-Snedecor; d) Gamma**

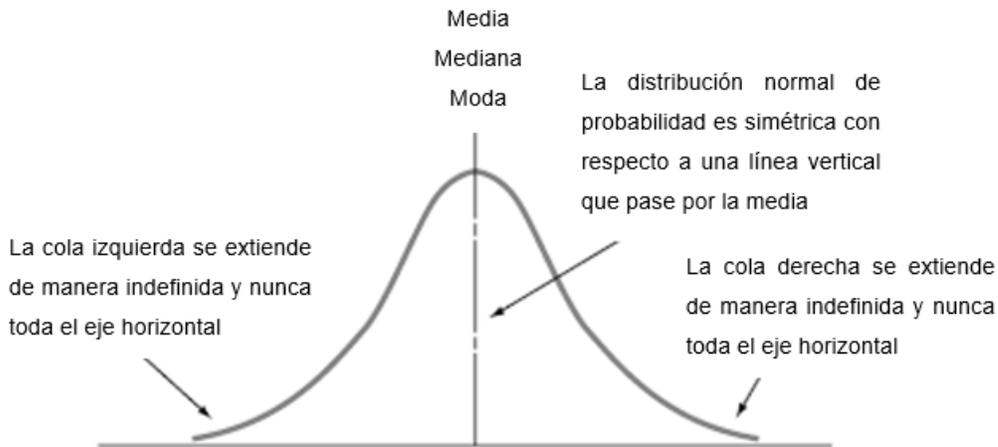


Fuente: Walpole et al. (2012). *Distribuciones continuas* (p. 202, 737, 739 y 741).

De acuerdo con Walpole, et al. (2012) la distribución continua más importante es la normal. Varios matemáticos contribuyeron a su desarrollo, pero el más importante fue Karl Gauss, por quien la distribución también recibe el nombre de Distribución gaussiana.

Su representación gráfica (figura 11), denominada como curva normal, tiene forma de campana, la cual describe de manera aproximada muchos fenómenos biométricos y procesos físicos, además de ser la distribución en la que se basan las inferencias para el muestreo probabilístico. Razones por las cuales Levin et al. (2004) se refieren a esta distribución como la más importante.

Figura 11. **Curva de distribución normal**



Fuente: Levin et al. (2004). *Curva de frecuencias para la distribución normal de probabilidad* (p. 209).

2.10. Distribuciones muestrales

El propósito del muestreo de poblaciones es obtener estimadores que se aproximen a los parámetros poblacionales (Walpole et al., 2012).

Según lo indican Sheaffer y Mendenhall (1997) las medidas numéricas usadas para describir las características que se esperan o expectativa de una población están dadas y/o asociadas al valor esperado o esperanza poblacional, dicho valor esperado de la población Y está dada por:

$$E(y) = \sum_y y p(y) \quad (\text{Ec. 5})$$

donde:

$Yp(y)$ = probabilidad asociada a la variable aleatoria y

El valor esperado es semejante al valor medio de todas las mediciones por lo que también es llamado media poblacional y se define como:

$$\mu = E(y) \quad (\text{Ec. 6})$$

dónde y es un valor individual seleccionado de la población al azar.

Alvarado (2007) explica que una característica en la mayoría de las poblaciones es que a medida que se incrementa el número de muestra, específicamente en el punto en el que se hace representativa y significativa, se observa que las frecuencias adquieren similaridad a la curva normal descrita por Gauss, tanto en forma como en las características que la describen.

La varianza se utiliza para cuantificar la distancia de los datos, esta medida no sigue un comportamiento normal si no que tiende a una distribución de forma X^2 y se define como:

“El valor esperado, del cuadrado de la desviación entre una medición y seleccionada aleatoriamente y su valor medio μ ” (Sheaffer y Mendenhall, 1997, p. 7).

La varianza de una población está dada por:

$$V(y) = E(y - \mu)^2 = \sum_y (y - \mu)^2 p(y) = \sigma^2 \quad (\text{Ec. 7})$$

Sheaffer y Mendenhall (1997) proponen los símbolos \bar{y} y S^2 para representar a los estimadores de la media y la varianza. Los modelos matemáticos que describen los estimadores están dados por:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (\text{Ec. 8})$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \quad (\text{Ec. 9})$$

dónde y_i es cada una de las observaciones y n es el tamaño de la muestra.

El muestreo probabilístico es realizado por la influencia de inconvenientes con los recursos, problemas técnicos, logísticos o cualquier otro motivo que impida el censo.

Como aporte a las consideraciones de Alvarado (2007) en relación al incremento de la muestra Sheaffer y Mendenhall (1997) afirman que también se reduce el error de muestreo que se define como la diferencia que existe entre el estimador y parámetro, una forma de cuantificarlo es por medio del límite para el error de estimación con $\alpha = 95 \%$ específicamente para la ecuación 10.

$$2\sqrt{\hat{V}(\bar{y})} = 2\sqrt{\frac{s^2}{n} \left(\frac{N-n}{N}\right)} \quad (\text{Ec. 10})$$

dónde:

S^2 = varianza de la muestra

n = número de muestra

N = total de la población

Sheaffer y Mendenhall (1997), Alvarado (2007) y Walpole et al. (2012) afirman que los estimadores son aproximados a los parámetros que se obtendrían al medir a todos los componentes de la población, ya que por medio del teorema del límite central es posible modelar cualquier distribución, media vez la muestra sea suficiente.

Debido al efecto del muestreo y la variabilidad entre los individuos existe el error de muestreo (ecuación 11) y es inherente a todas las estimaciones, este error indica la calidad del muestreo y no debe sobrepasar el error de estimación estipulado en la ecuación 4 (Scheaffer y Mendenhall, 1997).

$$\mathbf{Error\ de\ estimaci3n} = |\theta - \hat{\theta}| < \mathbf{B} \quad (\text{Ec. 11})$$

d3nde:

θ = estimador

$\hat{\theta}$ = parámetro

B = Límite para el error de estimación

2.11. Pruebas de bondad de ajuste

Son procedimientos utilizados para comparar la similitud de una distribución de frecuencias observada respecto a una teórica, la medida que será utilizada para la comparación de similitudes puede variar y es en ese último punto en donde se originan las distintas pruebas que se han establecido y se definen la potencia y características que deben ser analizadas para su selección (Webster, 2001).

Walpole et al. (2012) sugieren que toda vez que el investigador formule un cuestionamiento de probabilidades debe relacionarlo a la formulación de

hipótesis. Para comprobar el ajuste de una distribución muestral es necesaria la formulación de hipótesis.

Por consenso general se establece que la hipótesis nula (H_0) sugiere ajuste a la distribución teórica, por lo tanto, si existe similitud entre las frecuencias observadas y teóricas, se puede afirmar que la H_0 es verdadera (Webster, 2001).

2.11.1. Prueba X^2

Es una prueba no paramétrica útil para realizar pruebas de bondad de ajuste y de independencia.

Según Webster (2001) la prueba X^2 realiza el contraste de una hipótesis asociada a una distribución teórica, la cual se refiere a la distribución planteada como hipótesis y los datos observados obtenidos en el proceso de muestreo. La comparación es realizada por medio de la fórmula para la prueba de X^2 :

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (\text{Ec. 12})$$

dónde:

O_i = Frecuencia de los eventos observados en los datos muestrales

E_i = Frecuencia de los eventos esperados si la hipótesis nula es correcta

K = Número de categorías o clases

Los grados de libertad que serán utilizados en la distribución X^2 son:

$$gl = K-m-1 \quad (\text{Ec. 13})$$

dónde:

m = número de parámetros a estimar.

Webster (2001) sostiene que el numerador de la ecuación 12 mide la diferencia entre la probabilidad de eventos observados y esperados, razón por la cual, el estadístico incrementa proporcionalmente a la diferencia encontrada, situando la probabilidad en la región de rechazo de H_a .

Para seleccionar el valor crítico X^2 se utiliza la tabla 1 del anexo y si el valor elegido es mayor que el calculado se acepta H_0 .

La prueba X^2 se puede emplear para el contraste de variables continuas y discretas, debido a que para el ajuste se utilizan frecuencias y no eventos observados no es aplicable a poblaciones o muestras pequeñas.

2.11.2. Prueba de Kolmogorov-Smirnov

Es una prueba no paramétrica a la que se nombró en honor a los estadísticos que la desarrollaron A. N. Kolmogorov y N. V. Smirnov. Es una prueba utilizada para medir las diferencias entre una distribución de frecuencias de eventos observados y otra teórica. Esta prueba es de tipo libre, es decir que también puede ser utilizada para medir bondad de ajuste en distribuciones paramétricas y no paramétricas. Badii, Guillen, La, Araiza, Cerna, Valenzuela & Landeros (2012), mencionan otras pruebas de distribución libre como X^2 , tablas de contingencia, Wald-Wolfwitz, Cox, Cochran, Wilcoxon, Mann-Whitney entre otras.

Kolmogorov-Smirnov es una técnica válida para variables continuas y discretas, es aplicable a cualquier tamaño de muestra (Badii et al., 2012). De acuerdo con Márquez (2010) utiliza la probabilidad acumulada por lo que se considera más eficiente que la prueba X^2 .

Levin y Rubin (2004) consideran que la prueba es eficiente, sencilla y que el estadístico es de distribución libre porque se relaciona con el tamaño de la muestra, pero no con la probabilidad esperada.

El estadístico se define como una resta entre frecuencias y se selecciona la que sea mayor, tal como se describe en la ecuación 14 (Levin y Rubin, 2004).

$$D_n = \text{MAX} (F_e - F_o) \quad (\text{Ec. 14})$$

dónde:

F_e = frecuencias acumuladas relativas esperadas

F_o = frecuencias acumuladas observadas relativas

Para el cálculo del valor crítico D_n se utiliza la tabla 2 del anexo y se sustituye en la ecuación 15.

$$\text{Valor crítico para } D_n = \frac{1.63}{\sqrt{n}} \quad (\text{Ec. 15})$$

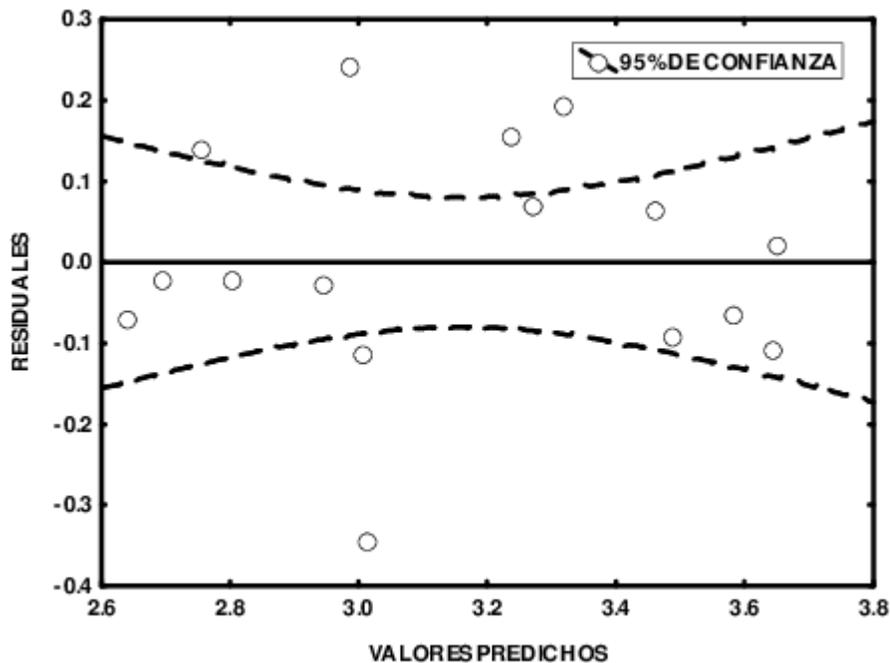
Si el valor de la tabla es mayor que el D_n obtenido a través de la ecuación 15 se acepta H_0 , que para este caso sugiere que la distribución teórica describe a la observada.

2.11.3. Jarque-Bera

Es una prueba utilizada para el contraste de normalidad. La prueba Jarque-Bera se basa en los conceptos de regresión lineal y compara la simetría y curtosis de la densidad teórica respecto a la empírica, para lo que hace uso de los valores predichos de la estimación (Jardón, 2007; Vela, 2010).

Una distribución que presenta similitud respecto a la normal presenta mayor frecuencia en la región cercana a cero y se hace más dispersa mientras las frecuencias se hacen menores (figura 12), los valores muy dispersos pueden sugerir datos atípicos (Vela, 2010).

Figura 12. **Gráfico de predichos y residuales**



Fuente: Gómez J. (2015). *Notas sobre el análisis de regresión lineal múltiple* (p. 69).

Jardón (2007) expresa que para realizar el contraste de hipótesis se deben calcular los residuos estandarizados, posteriormente se ordenan de menor a mayor con el objetivo de hacer coincidir el cuantil de orden t/T y se calcula el valor crítico de Jarque-Bera tal como se expresa en la ecuación 16.

$$JB = N * \left\{ \frac{m_3^2}{6} + \left(\frac{m_4 - 3}{24} \right)^2 \right\} \quad (\text{Ec. 16})$$

dónde:

N = número de observaciones

m_3 = medida de asimetría

m_4 = medida de curtosis

La prueba compara la asimetría y la curtosis bajo el supuesto de que en una curva normal ambos son igual a cero, de tal forma que si alguna de estos dos valores es muy diferente de cero no existe normalidad (Vela, 2010; Jardón 2007).

Para el contraste de la $H_0 = m_3 = m_4 = 0$, se utiliza la distribución de probabilidad X^2 y dos grados de libertad (Jardón, 2007), tal como se expresa en la ecuación 17:

$$|JB| > X_{2,\alpha} \quad (\text{Ec. 17})$$

dónde:

$|JB|$ = valor crítico calculado de Jarque-bera

$X_{2,\alpha}$ = Probabilidad X^2 tabulada

2.12. Métodos paramétricos para el contraste de hipótesis

Una hipótesis es una conjetura que se realiza cuando se quiere medir probabilidades, en términos estadísticos existe una probabilidad de ocurrencia y de no ocurrencia y para medirla se plantean H_0 y la H_a (Walpole et al., 2012).

Para el caso de la prueba de hipótesis de dos o más muestras describen los siguientes supuestos:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

El parámetro o estimador utilizado para la medición de hipótesis depende de los objetivos del investigador, para el caso de la media se utiliza la distribución normal y el estadístico utilizado para la cuantificación del área bajo la curva Z .

$$Z = \frac{\bar{X} - \mu_{H_0}}{\sigma_{\bar{X}}} \quad (\text{Ec. 18})$$

dónde:

\bar{X} = media de la muestra

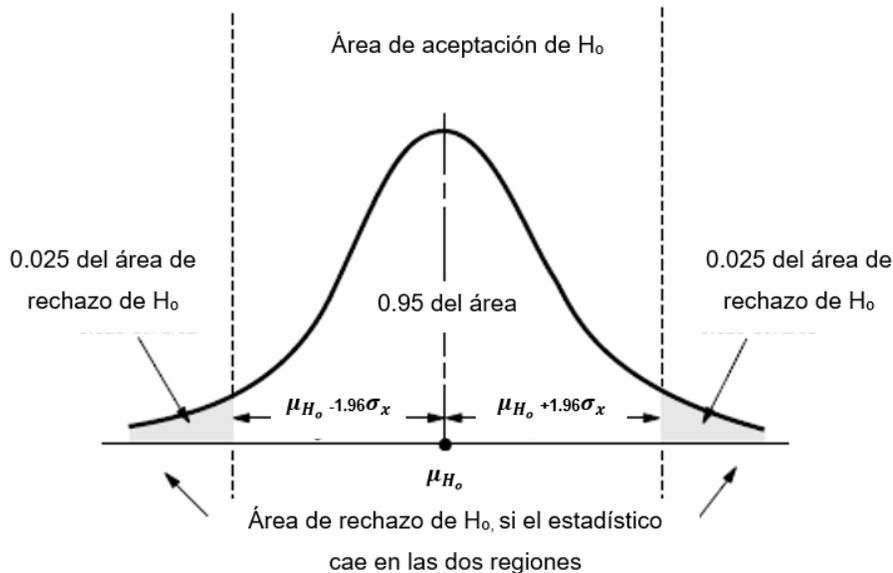
μ_{H_0} = media hipotética

$\sigma_{\bar{X}}$ = desviación estándar de la muestra

El propósito de la prueba de hipótesis no es cuestionar el valor calculado del estadístico de la muestra, sino hacer un juicio respecto a la diferencia entre ese estadístico y un parámetro hipotético de la población (...) el utilizado para confirmar si se acepta o se rechaza la hipótesis nula se conoce como nivel de significancia. (Levin y Rubin, 2004, p. 325).

El nivel de significancia (α) representa la probabilidad de rechazo de H_0 , según el grado de error que el investigador esté dispuesto a aceptar en sus resultados se selecciona entre 1 %, 5 % y 10 %, el valor se ubica en las colas del modelo según se haya planteado la hipótesis, para el planteamiento de hipótesis descrito al inicio de esta sección la figura 13 ejemplifica la forma en que se distribuyen las probabilidades (Levin y Rubin, 2004; Walpole et al., 2012).

Figura 13. **Área de aceptación y rechazo para H_0 (α 5 %)**



Fuente Levin y Rubin (2004). *Interpretación del nivel de significancia* (p. 326).

Utilizar la información del gráfico y del estadístico Z es crucial en el análisis de conclusiones debido a que la ubicación de Z indica si se aceptan o rechazan las hipótesis, para transformar el valor de Z en área bajo la curva se usa como referencia la tabla 3 de los anexos y posterior a ello se relaciona su ubicación en el modelo para conocer la probabilidad del evento.

2.13. Prueba de Grubbs

Un valor atípico es aquel que parece no tener correspondencia con relación al conjunto de datos al que pertenece. Según lo indica Mejía (2007) retirar un valor atípico debe estar respaldado por una prueba estadística que sugiera un nivel de confianza aceptable.

La prueba de Grubbs calcula diferencias absolutas entre la media del conjunto de datos y el valor analizado, lo divide dentro de la desviación estándar del mismo conjunto de datos, tal como lo refiere Mejía (2007).

$$T_{calculado} = |X_0 - \bar{X}|/S \quad (\text{Ec. 19})$$

Con base en el tamaño del conjunto evaluado y el nivel de significancia se calcula el $T_{teórico}$ y si el $T_{calculado} < T_{calculado}$, entonces se concluye que el valor analizado es atípico.

3. PRESENTACIÓN DE RESULTADOS

3.1. Distribución de probabilidad de ajuste al peso de racimos de palma de aceite

Se midió el peso a los individuos de 15 poblaciones contenidas en unidades de transporte, la cantidad de racimos contenidos en cada unidad se describe en la tabla 5.

Tabla V. **Número de racimos por medición**

Medición	Racimos pesados
1	403
2	366
3	167
4	287
5	100
6	200
7	317
8	353
9	311
10	541
11	471
12	413
13	392
14	469
15	723
Total, general	5513

Fuente: elaboración propia.

Para comprobar el ajuste del peso de racimos se realizaron varias pruebas de normalidad, entre ellas la prueba de X^2 , Kolmogorov-Smirnov y Jarque-Bera.

Pedrosa, Juarros, Robles, Basteiro y García (2014) consideran que la prueba de Shapiro Wilk, no es representativa para evaluaciones superiores a 50 unidades, en congruencia con lo representado en la tabla V no es una prueba representativa para el desarrollo del presente estudio.

Al tomar en cuenta las recomendaciones de Pedrosa et al. (2004) se prosiguió con los procedimientos de las pruebas de X^2 y Kolmogorov-Smirnov, ambas pruebas con base en las consideraciones necesarias para un grado de significancia de 5 %.

Los resultados de ambas pruebas sugieren que las distribuciones de frecuencias evaluadas no se ajustan a la distribución propuesta por Gauss. Es importante describir que por medio de la de Kolmogorov-Smirnov se obtuvo ajuste en algunas de las poblaciones evaluadas situación que contrasta con los resultados de la prueba X^2 en dónde no se logró ajuste en ninguna de ellas.

Pedrosa et al. (2014) pueden colaborar para comprender esta falta de ajuste, ya que expresan que la prueba de K-S es una prueba clásica con tendencias conservadoras, lo que implica la probabilidad de aceptar la hipótesis de normalidad en la mayoría de las ocasiones. Además de que la prueba es más sensible en el centro que en las colas y que debido a que la media y la varianza son calculadas en base a la muestra, disminuye la validez de la región crítica.

Respecto a la mayor sensibilidad de la prueba K-S la forma de los histogramas ilustrados en el anexo 2 pueden apoyar los argumentos

anteriormente descritos, ya que la falta de ajuste de distribuciones se observa en las colas, es decir extremos mínimos y máximos del peso.

Existen otros tipos de pruebas con mayor potencia estadística, entre las que se puede destacar a la metodología de Jarque-Bera (Pedrosa et al., 2014).

De tal forma los datos fueron nuevamente analizados a través de la prueba Jarque-Bera. En la tabla VI se aprecia que a pesar de que se obtuvo un mayor porcentaje de poblaciones ajustadas, no se logra el ajuste para todo el conjunto de poblaciones.

Tabla VI. **Valores críticos para prueba de Jarque-Bera**

Medición	p-valor Jarque-Bera
1	0.2437
2	0.6048
3	0.06472
4	0.7265
5	0.02407
6	1.02E-06
7	0.0009625
8	0.05679
9	0.4656
10	6.82E-05
11	6.97E-05
12	2.20E-16
13	2.20E-16
14	0.000339
15	0.04298

Fuente: elaboración propia.

Para cumplir con el supuesto de normalidad, se decidió aplicar procedimientos de ajuste a los valores de peso obtenidos en la fase de campo. El método seleccionado para el ajuste o transformación fue la metodología de BoxCox. Este tipo de transformación fue seleccionada por su compatibilidad en criterios de comparación, Jarque-Bera utiliza criterios de la regresión lineal para sesgo y curtosis, y BoxCox es aplicada en la corrección del sesgo en los errores de estimación (Montero R, 2016).

Posterior a la transformación de datos se realizó nuevamente la metodología de Jarque-Bera. Los valores de λ obtenidos a través de la prueba BoxCox y los nuevos p-valores correspondientes los nuevos cálculos de la prueba Jarque-Bera se detallan en la tabla VII.

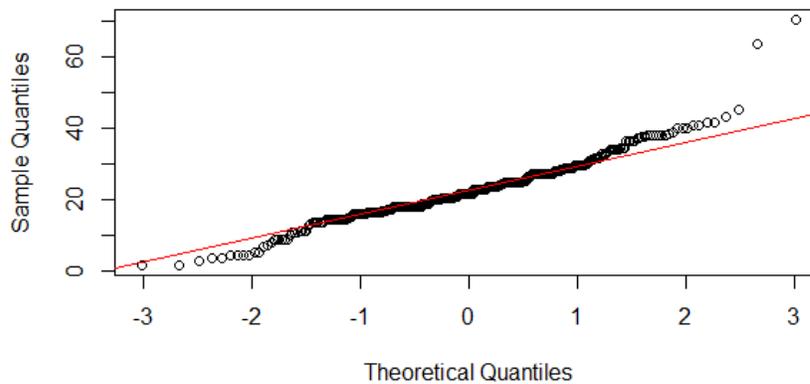
Tabla VII. λ para transformación y p-valor para bondad de ajuste

Medición	λ BoxCox	p-valor Jarque-Bera
1	0.8	0.8447
2	0.95	0.7364
3	0.4	0.2311
4	0.95	0.7575
5	0.4	0.9167
6	0.4	0.7405
7	0.5	0.9756
8	0.6	0.1844
9	0.85	0.9633
10	0.75	0.1712
11	0.8	0.1468
12	0.55	6.46E-08
13	0.9	0.2157
14	0.85	0.01352
15	0.8	0.4258

Fuente: elaboración propia.

Es importante resaltar que las poblaciones 12 y 14 son consistentemente no normales. Para incluir en el análisis paramétrico a la población 13, se le aplicó el criterio de Grubbs, la representación gráfica de atípicos se representa en la figura 14.

Figura 14. **qqplot medición 13, identificación de puntos atípicos**



Fuente: elaboración propia, utilizando con RStudio.

Para el caso de las poblaciones 12 y 14 no se tomó este criterio debido a que son evidentemente no normales, tal como se presenta en los histogramas y gráficos qq-plot de los apéndices 2 y 3. Los p-valores con los que se aceptó que los dos valores son valores atípicos fueron de $9.34e-07$ y $3.287e-05$, para una hipótesis alterna que propone que son datos atípicos.

En la serie de datos transformada por la metodología BoxCox se evaluaron los valores de curtosis, los coeficientes obtenidos para las poblaciones que no se ajustaron a la curva normal fueron de 1.38 y 0.64, los cuales sugieren que dichas distribuciones poseen colas más largas o pesadas en comparación a las colas de una distribución normal.

3.2. Tamaño muestral requerido para la medición de la calidad de fruta

La determinación del número de muestra se realizó con la fórmula de Scheaffer y Mendenhall (1997), $\alpha = 5\%$ y $B = 3$ kg, debido a la poca coincidencia con la curva de Gauss las poblaciones 12 y 14 se omitieron del análisis (tabla VIII).

Tabla VIII. **Tamaño de muestra por medición**

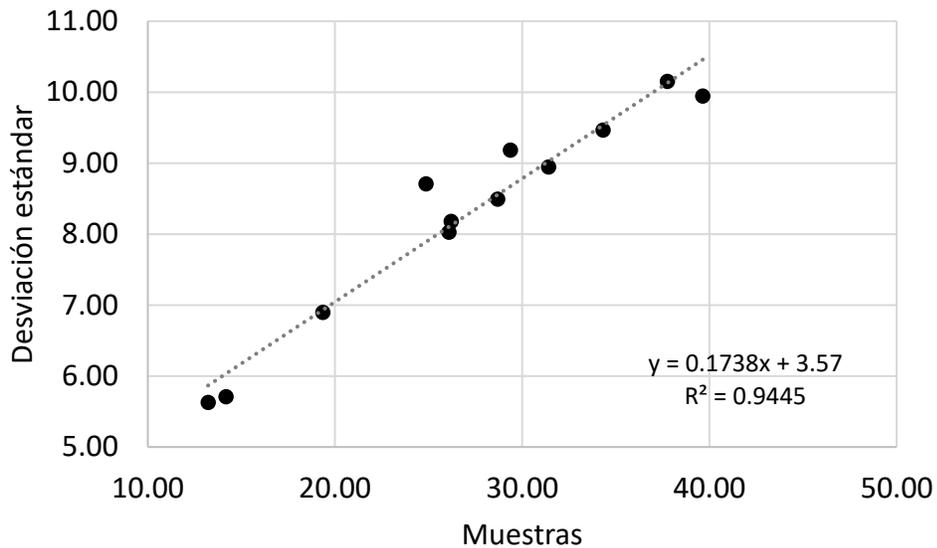
Medición	Número de muestra	Media de peso (kg)	Desviación estándar
1	34.32	22.51	9.47
2	26.20	23.45	8.18
3	29.37	28.00	9.18
4	31.41	27.84	8.95
5	24.86	22.13	8.71
6	13.22	17.44	5.63
7	37.75	28.08	10.15
8	26.09	27.37	8.02
8	17.39	18.91	6.41
10	14.17	15.97	5.71
11	39.65	23.47	9.94
13	28.68	22.90	8.49
15	19.34	16.27	6.90
Promedio	26.34	22.64	8.13

Fuente: elaboración propia.

Se requieren entre 13.22 y 39.65 racimos para el muestreo representativo y significativo al 5 % según las poblaciones descritas en el presente documento. La media para el tamaño de la muestra es de 26.34 racimos ~ 27 racimos/unidad de muestreo.

Nótese que las mediciones con mayor desviación estándar son las que sugieren un mayor número de muestra, congruente con la correlación representada en la figura 15.

Figura 15. **Tamaño muestral y desviación estándar poblacional**



Fuente: elaboración propia.

Para comprobar si $n = 27$ racimos por unidad de transporte es un número representativo de la unidad de muestreo, se procedió a realizar un proceso de remuestreo aleatorio con los datos de cada unidad de transporte (cinco repeticiones por medición). Se utilizó la metodología de muestreo sin reemplazo, ya que en la práctica los evaluadores de calidad pintan los racimos muestreados para seleccionarlos una sola vez. Para brindar mayor sustento a la calidad del muestreo se complementó la información de los remuestreos con el cálculo del error de estimación (tabla IX).

Tabla IX. **Comparación de coeficientes de variación**

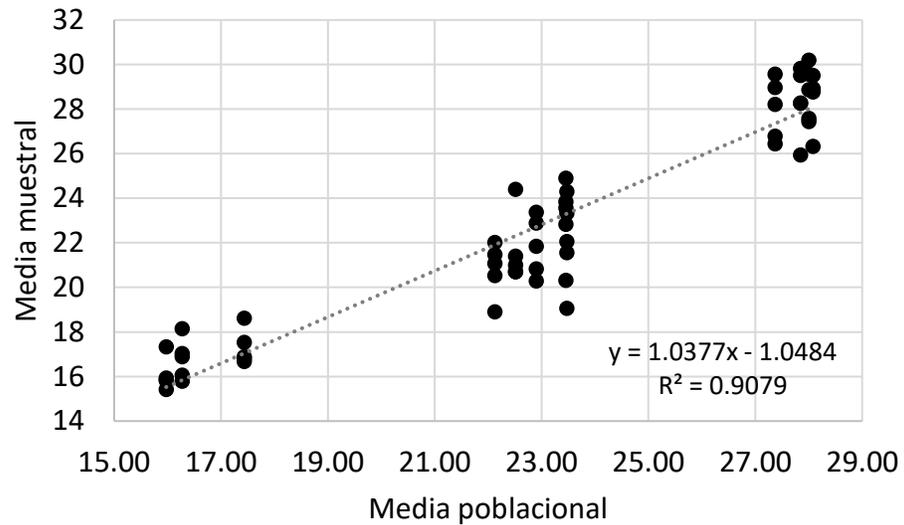
Muestreo	CV población	Promedio de CV muestra	Promedio de error de estimación (kg)
1	2.38	2.44	1.62
2	2.87	3.07	1.15
3	3.05	3.03	1.12
4	3.11	3.48	1.28
5	2.54	2.64	1.32
6	3.10	3.07	0.64
7	2.77	2.90	1.11
8	3.41	3.55	1.23
9	2.95	2.88	0.87
10	2.80	2.93	0.44
11	2.36	2.59	1.74
13	2.70	2.87	1.24
15	2.36	2.69	0.79
promedio	2.80	2.93	1.12

Fuente: elaboración propia.

Los coeficientes de variación de la muestra y la población son similares y el promedio del error de estimación de las cinco repeticiones es menor al error de estimación utilizado en la fórmula $B = 3$ kg por lo que según lo expresan Scheaffer y Mendenhall (1997) las estimaciones son significativas al 5 %.

Por medio del gráfico de regresión representado en la figura 16, en dónde se analizó la media poblacional y la media muestral, se encontró que existe alta correspondencia entre ambas mediciones ($R^2 = 0.9079$).

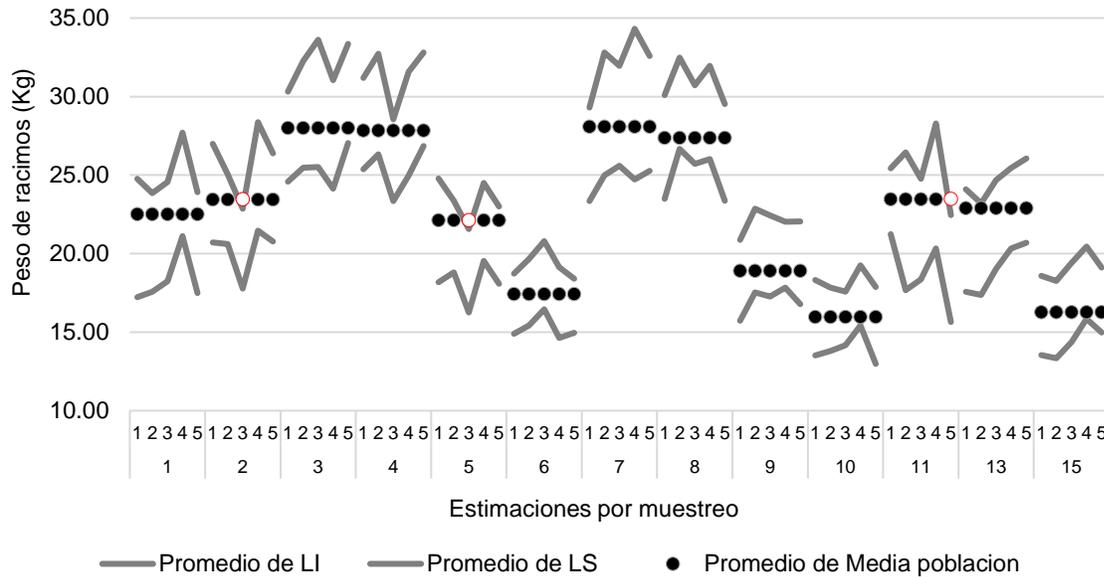
Figura 16. **Correlación entre la media poblacional y las medias de muestreo**



. Fuente: elaboración propia.

Con la misma información obtenida en el proceso de remuestreo, se calculó el límite de estimación para el error. La figura 17 contiene remarcados los tres puntos que no están contenidos dentro de los límites, las tres medias que salen de estos límites equivalen al 4.62 % de los datos graficados, valor que coincide y no sobrepasa al nivel de significancia.

Figura 17. Intervalo para el error de estimación



Fuente: elaboración propia.

3.3. Comparación de dos métodos de muestreo para la selección de la muestra de racimos

Tal como se describió en la parte metodológica se evaluaron dos métodos de selección para la muestra. El primer método evaluado fue el método de la cuerda, el cual es el que se adoptó en Guatemala como resultado de los estudios realizados en Colombia por García y Yáñez (2000). La segunda propuesta es una variación del método de la cuerda que consistió en dividir la unidad de transporte en cuatro cuadrantes iguales, seleccionar una muestra aleatoria dentro de cada cuadrante hasta completar una muestra de 30 racimos, seleccionados igualmente por medio de los nudos equidistantes de la cuerda.

Para comparar las metodologías propuestas se procedió a realizar pruebas t Student para muestras independientes. Según los resultados presentados en la tabla 10 y con $\alpha = 5 \%$, no existe diferencia estadística entre las medias de ambas metodologías de selección.

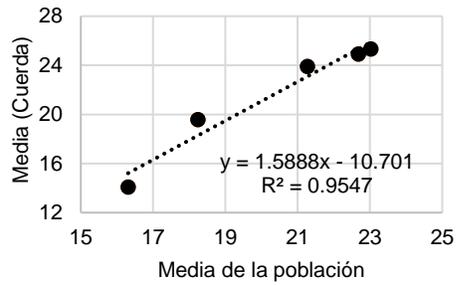
Tabla X. **Comparación de medias para métodos de selección de muestra**

	Repetición	cuerda	cuadrante
Media	1	23.89	21.24
Varianza		26.68	38.73
p-valor		0.07776	
Media	2	25.32	23.03
Varianza		70.77	40.44
p-valor		0.2395	
Media	3	19.58	18.55
Varianza		39.12	29.91
p-valor		0.4998	
Media	4	24.91	25.71
Varianza		79.4	52.66
p-valor		0.7033	
Media	5	14.09	15.79
Varianza		28.58	15.83
p-valor		0.1684	

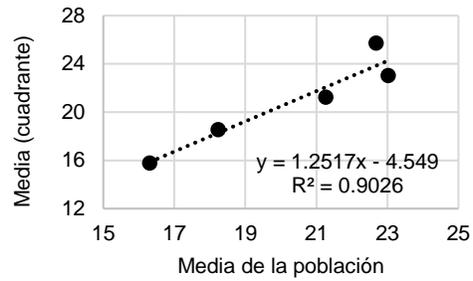
Fuente: elaboración propia.

Así mismo, para conocer la representatividad del método de muestreo se realizaron correlaciones (figura 18) entre la media de las muestras obtenidas por medio del método convencional y del cuadrante.

Figura 18. **Correlación de medias para métodos de muestreo**



a)



b)

Fuente: elaboración propia.

Los R^2 son lo suficientemente altos como para argumentar una buena correlación y son consistentes con los resultados de la figura 18.

4. DISCUSIÓN DE RESULTADOS

- Análisis de la distribución de probabilidad a la que se ajusta el peso de racimos

En el análisis exploratorio de los datos se identificó la presencia de colas gruesas en las distribuciones de frecuencia de algunas poblaciones, razón por la cual previo a la transformación de datos BoxCox no se obtuvo el ajuste esperado.

A pesar de la aplicación de los métodos sugeridos por BoxCox no se obtuvo ajuste normal en todas las poblaciones, razón por la cual dos de ellas fueron excluidas del análisis paramétrico de la información.

Según Ormerod (2010) una característica de la distribución normal es que la frecuencia relativa de las colas es corta, pero para el caso de los pesos de racimos contenidos en las unidades de transporte 12 y 14 es un fenómeno al que se le conoce como colas gruesas.

Murillo (2009) resalta que las distribuciones que pueden describir el comportamiento de una distribución de colas pesadas o gruesas son las curvas de Frechet, t Student, Pareto, LogGamma o Cauchy.

La falta de normalidad en las curvas de frecuencia del peso de la fruta de palma puede ser consecuencia de factores que influyan en el proceso de llenado de los racimos o en el desprendimiento de frutos. Entre estos factores se pueden mencionar los fisiológicos, climáticos y de manejo agronómico.

Para ampliar el análisis se considera que cualquier tipo de estrés al que la planta se encuentre expuesta puede desencadenar reacciones fisiológicas que afecten el peso y número de racimos. Las condiciones meteorológicas, principalmente la radiación, temperatura, humedad relativa y precipitación poseen influencia sobre la maduración de los frutos, la germinación del polen y sobre agentes externos como los polinizadores. La formación, maduración y calidad de los racimos está directamente influenciada por estos factores, de los cuales la precipitación y la temperatura también tienen influencia sobre el desprendimiento de frutos.

Por otra parte, el manejo agronómico de las plantaciones constituye otro factor importante, que además es sujeto de cambio y mejora. Durante el desarrollo del estudio se eliminó el factor de desprendimiento de frutos producido por la cosecha fuera de ciclo, ya que los racimos fueron cosechados dentro de los parámetros de ciclo de cosecha. El factor que no se tomó en cuenta fue el tiempo de residencia de la fruta en campo posterior a la cosecha y previa a la carga en las unidades de transporte. Según los registros la población 14 y parte de la fruta de la población 12 permanecieron más de 24 hrs en campo previo a la carga y peso. El factor no considerado también propicia el desprendimiento de la fruta y es esta última condición la que influye en la discrepancia respecto al modelo Gaussiano.

Los muestreos fueron seleccionados al azar y contienen unidades con fruta mezclada entre lotes y variedades, lo cual indica que esta heterogeneidad no constituye un factor para la pérdida de la normalidad, tanto como lo constituye el desprendimiento de fruto, después de la cosecha.

- Interpretación del tamaño muestral requerido para la medición de la calidad de fruta

Los resultados de los muestreos en los que se mezcló fruta de dos lotes son más variables, por consiguiente, el número de muestra calculado es mayor en comparación a los lotes que poseían fruta procedente de un solo lote. En promedio, es necesario extraer una muestra de 26.39 ~ 27 racimos por unidad de transporte.

El error de estimación entre la media de la muestra y la de la población fue de 1.2 kg, dato inferior a 3 kg, que fue el error de estimación utilizado en la fórmula para la determinación del número de muestra, de tal forma que es válido deducir la representatividad del estimador obtenido con 27 racimos. Aunado a los criterios anteriores también se comparó el nivel de correspondencia entre el estimador y el parámetro, calculado en $R^2 = 0.9079$, lo que sugiere a su vez, un bajo error de estimación.

Scheaffer y Mendenhall (1997) califican a la media muestral como un estimador insesgado, para el caso del presente estudio el estimador para el número de racimos necesario en el muestreo de calidad de fruta es de 27 racimos, con un intervalo de error de ± 2.85 .

Los intervalos de confianza fueron probados para verificar que se cumpliera con el supuesto del grado de confianza al 95 %, según las 65 muestras extraídas, el 4.62 % de los datos no están contenidos en el intervalo para el error de estimación, lo que es congruente con el grado de confianza planteado en la fórmula de Scheaffer y Mendenhall (1997).

La metodología de García y Yáñez (2000) propone la extracción de 30 racimos, según los resultados expresados anteriormente se considera acertado mantener el número de muestra recomendado por los colombianos.

- Análisis comparativo de dos métodos de muestreo para la selección de la muestra de racimos

Guatemala carece de estudios que propongan nuevas metodologías de muestreo de racimos. El método tradicional de la cuerda (descrito en los manuales de procedimientos de la operación agroindustrial de palma de aceite) y una variante del mismo método que se denominó método de los cuadrantes fueron evaluados para determinar su conveniencia.

Se realizaron 5 muestreos de tamaño $n = 30$ y sin reemplazo para cada método. Al hacer uso de los métodos propuestos por la prueba *t Student* no se encontraron diferencias significativas entre las medias obtenidas con ambos métodos.

La variación de las poblaciones está vinculada a la precisión de la estimación y a la desviación del peso inherente a la genética de la planta y a la influencia de factores exógenos que afectan la formación y desarrollo de los racimos. Las estimaciones realizadas con ambos métodos fueron equivalentes entre sí y se obtuvo una precisión aceptable para ambos métodos. Al comparar las varianzas y la regresión de medias estimadas respecto sus respectivos parámetros, ambos métodos sugieren una variabilidad similar.

La interpretación de los resultados sugiere que ambos métodos son válidos, toda vez se tomen en cuenta las recomendaciones de tamaño de n y los criterios de selección muestral.

Los resultados del estudio plantean bases estadísticas que dan sustento a las decisiones que se toman con base a la práctica de muestreo de calidad de fruta, ya que describe los fundamentos de inferencia estadística sobre los que se basó el diseño para el muestreo de racimos, con una significancia de 5 % y representativo para región palmera de Fray Bartolomé de las Casas.

CONCLUSIONES

1. El diseño del método de muestreo para la medición de calidad de fruta propuesto en el contenido del documento es estadísticamente significativo y representativo para las condiciones de Fray Bartolomé de las Casas.
2. La distribución de probabilidad a la que se ajusta el peso de racimos de palma de aceite es normal. Los factores ligados a la génesis del peso de racimos y el desprendimiento de frutos favorecen el fenómeno de colas gruesas, que distorsiona el ajuste respecto a la curva normal.
3. El tamaño muestral requerido para la medición de calidad de fruta en palma de aceite, con base en métodos descritos en el presente documento es de 26.34~27 racimos con un intervalo de error de ± 2.85 .
4. El contraste de hipótesis de los métodos de la cuerda y los cuadrantes ($\alpha = 0.05$) indica que no evidencia suficiente para sugerir la zona de rechazo de H_0 , que existe igualdad de medias y que ambas son representativas respecto a la media de la población.

RECOMENDACIONES

1. Para los departamentos de investigación de las empresas palmeras es importante profundizar en el fenómeno de las colas gruesas, las curvas de probabilidad asociadas a él y determinar con mayor precisión los factores de influencia, para brindar recomendaciones en campo encaminadas al buen manejo de las plantaciones y para que el muestreo de racimos en planta de beneficio sea lo más representativo posible.
2. Las plantas de beneficio de palma de aceite de otras regiones con características climáticas y de manejo agrícola diferentes a las referidas para Fray Bartolomé de las Casas, podrían validar los procedimientos planteados en el presente documento y ajustarlos a sus propias condiciones.

REFERENCIAS

1. Abdul, R.J., Zulkifli, Ab., Rahman y Salleh, N.M., 2009. *Curso de competencia en cultivo de racimos de palma*. Malaysian Palm Oil Board (MPOB), Selangor.
2. Alvarado, H. (2007). *Significados institucionales y personales del teorema central del límite en la enseñanza de estadística en ingeniería* (Tesis doctoral). Universidad de Granada, Granada.
3. Azli A, R., Shahril, K., Nor Jayati, S.y Sukanur C. A. (2009). *Novel Method of grading fresh fruit bunches (FFB) of oil palm fruit*. CSSR, 8(1). 1-9.
4. Badii, M.H., Guillen, A., Araiza, L.A., Cerna, E., Valenzuela, J. y Landeros, J., (2012). *Métodos no paramétricos de uso común*. International Journal of Good Conscience, 7(1). 126-149.
5. Bracamonte, E. (2018). *Aproximación a la distribución normal de los registros de humedad relativa en la ciudad de Guatemala durante el período 1978 a 1983, utilizando la prueba de Kolmogorov-Smirnov*. Revista de la Escuela de Estudios de Posgrado, 9(1). 30-34.
6. Central American Business Intelligence (CABI), 2016. *Impactos Socioeconómicos sobre el cultivo de palma de aceite en Guatemala*. P 34.

7. Córdova, M. (2003). *Estadística descriptiva e inferencial*. Lima, Perú. Moshera S.R.L.
8. Corley, R. y Tinker, P. (2003). *La palma de aceite*. Bogotá, Colombia. Fedepalma.
9. Días, S., Ojeda, M., y Valderrábano, E. (2016). *Metodología de muestreo de poblaciones finitas para aplicaciones en encuestas*. México, Imaginaria editores.
10. Eroles, R. (2016). *Calidad de cosecha*. Revista La palma, 5(1), 1-14.
11. Fairhurst, T. y Härdter, R. (2003). *Palma de aceite: Manejo para rendimientos altos y sostenibles*. International Plant Nutrition Institute & International Potash Institute.
12. Forumo, P. y Mitchell, T. (2017). *Caracterización de la expansión de la palma de aceite para uso comercial en América Latina: Cambio en el uso del suelo y comercialización*. Revista Palmas. 38(2), 27-48.
13. García, J. A. y Yáñez, E. E. (2000). *Aplicación de la metodología alterna para análisis de racimos y muestreo de racimos en tolva*. Palmas, 21(1), 303-310.
14. Gómez, J. (2015). *Elementos de farmacología electrónica molecular*. Chile. Edición electrónica.
15. Gremial de palmicultores de Guatemala (GREPALMA), 2017. *I Anuario Estadístico 2016-2017*. Guatemala. Editorial Serviprensa.

16. Hazir M., Shariff A. Amiruddin M. (2012). *Determination of oil palm fresh fruit brunch ripness-Based on flavonoids and anthocyanin content*. Industrial Corps and Products, 36(1), 466-475.
17. Jardón, C. (2007). *La normalidad de las perturbaciones*. Granada, España: Uvigo. Recuperado de: <http://cjardon.webs.uvigo.es/Transparencias/Unidad2.pdf>.
18. Levin, R. y Rubin, D. (2004). *Estadística para administración y economía*. México. Pearson Educación.
19. Lohr, S. (1999). *Muestreo: Diseño y Análisis*. International Thomson Editores.
20. Maibaum, G. (1988). *Teoría de probabilidades y estadística matemática*. Recuperado de <https://es.scribd.com/document/382212506/Gert-Maibaum-Teoria-de-probabilidades-y-estadistica-matematica-pdf>.
21. Márquez, C., (2010). *Prueba de bondad de ajuste*. Universidad Nacional de Colombia. Recuperado de <https://carlosmarquez.files.wordpress.com/2012/02/prueba-de-bondad-de-ajuste.pdf>.
22. Meftha, S., Abdul R., Mohd h., Saared O. y Eshanta O. (2008). *Oil palm fruit brunch grading system using red, green and blue digital number*. Journal of Applied Sciences, 8(8), 1444-1452.
23. Mejía, G. (2007). *Prueba de Grubbs*. Recuperado de <https://www.academia.edu>.

24. Mohd, M. (2002). *Tasa de extracción de aceite de palma (TEA) de Malasia: Una respuesta de factores de mercado y administración*. Revista Palmas. 23(1). 41-50.
25. Montero, R. (2016). *Modelos de regresión múltiple*. Recuperado de https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf
26. The Malaysian Palm Oil Board (MPOB), 2006. *Manual de procesamiento de palma de aceite*, 3 ed. Malasian Palm Oil Board. (MPOB), Selangor.
27. Murillo, G. (2009). *La teoría de valor extremo y el riesgo operacional: una aplicación en una entidad financiera*. Ingenierías Universidad de Medellín, 8(15), 59-70.
28. Ormerod, P. (2010). *La Crisis actual y la culpabilidad de la teoría macroeconómica*. Revista de institucional, 12(22), 111-1128.
29. Pedrosa, I., Juarros, J., Robles, A., Basteiro, J. y García, E. (2014). *Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar?*. Universitas Psychologica, 14(1) 15-24.
30. Rashid, A. (2016). *Clasificación de Racimos de fruta fresca en la planta de beneficio de aceite de palma utilizando técnicas y tecnologías avanzadas*. Palmas, 37(2), 11-18.
31. Rodríguez, J., Ferreras, L y Núñez, A. (1978). *Inferencia estadística, niveles de precisión y diseño muestral*. Reis, 54(9) 139-162.

32. Romero, M. (2016). *Pruebas de bondad de ajuste a una distribución normal. Metodología de la Investigación*, 6(3), 105-114.
33. Scheaffer, R. y Mendenhall, W. (1997). *Elementos de muestreo*. México. Grupo editorial Iberoamérica.
34. Sinsinaj, J. (2007). *Costos y rentabilidad de unidades agrícolas (producción de cardamomo)*. Tesis Contador público y auditor. Guatemala, USAC, 281 p.
35. Sandoval, W. (2007). *Estudio de mercado de pimienta negra (Piper nigrum) y contribución a mejorar los sistemas de producción agrícola y forestal de la comunidad nueva libertad del municipio de Fray Bartolomé de las Casas, Alta Verapaz*. Tesis Ing. Agr. Guatemala, USAC, 84 p.
36. Taillez, B., Siaka, M., Bonny, C.P. y Jacquemard, J.C. (1996). *La maduración de los racimos de palma y los criterios de cosecha en Palmindustrie (Costa de Marfil)*. *Revista Palmas*. 17(1). 29-37.
37. The Thai Agricultural Standard. (2008). *Oil Palm Brunch*. p 5.
38. The Plant List. 2014. *A workinglist of all plant especies* (en línea). London. Consultado 3 mar 2014. Disponible en <http://www.theplantlist.org/tpl1.1/record/kew-254945>
39. Toong, T.H., Yang, T.S. (1993). *Normas de cosecha y control de calidad para una mayor productividad de la palma africana*. *Palmas*. 14(2). 63-70.

40. Vecco, D., Pinedo, R. y Fernández, M. (2015). *Métodos de muestreo para Pseudophillothrips sp. (tubulífera: Phlaeothripidae) y sus enemigos naturales: I. Muestreo secuencial enumerativo*. *Protección vegetal*, 30(1) 6-13.

41. Vela, F. (2010). *Normalidad de los errores*. D. F., México: mregresion. Recuperado de: <https://mregresion.files.wordpress.com/2011/10/normalidad.pdf>

42. Walpole, R., Myers, R., Myers, S. y Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias*. México. Pearson Educación.

43. Webster, A. (2012). *Estadística aplicada a los negocios*. Colombia. McGraw-Hill.

APÉNDICES

Apéndice 1. Formato para registro de información de peso y número de racimos en una unidad de transporte

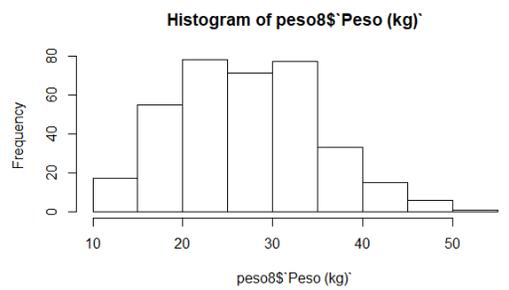
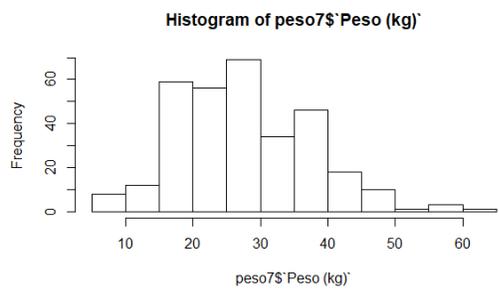
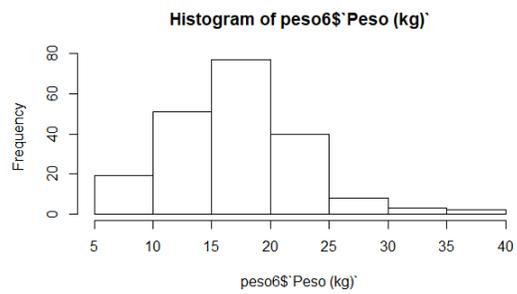
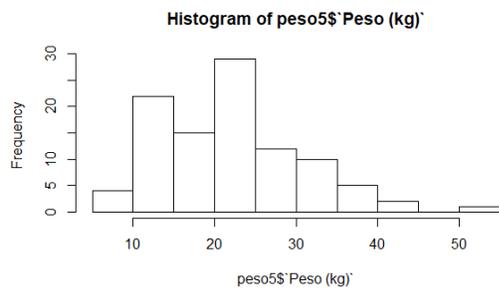
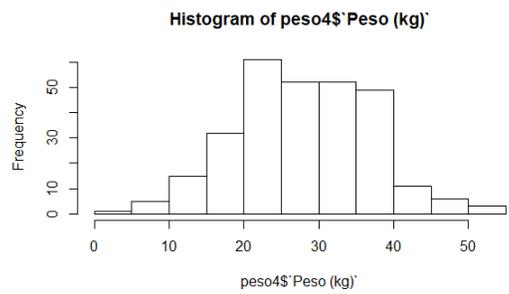
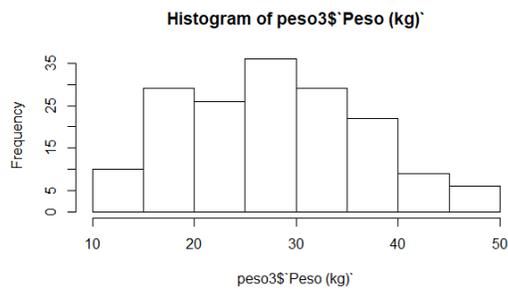
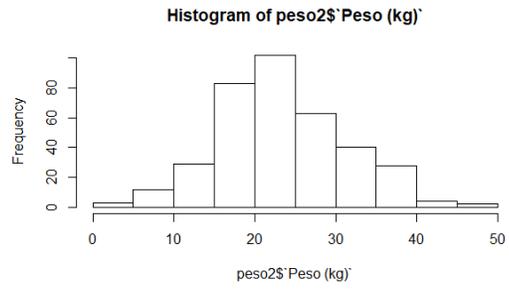
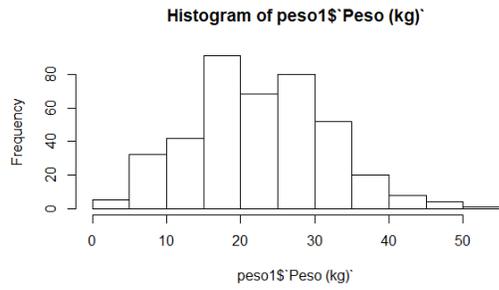
Fecha: _____ Finca: _____

Encargado: _____ Placa
camión: _____

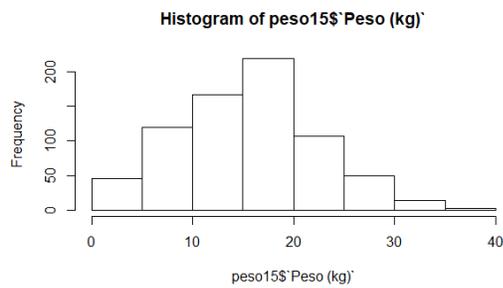
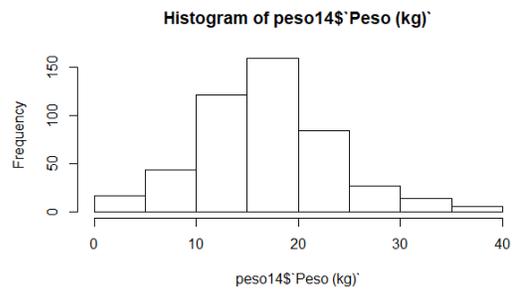
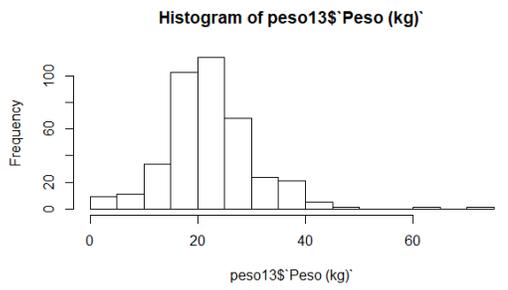
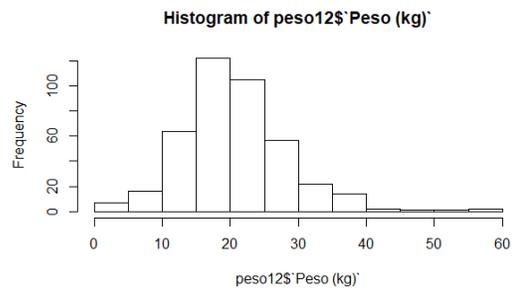
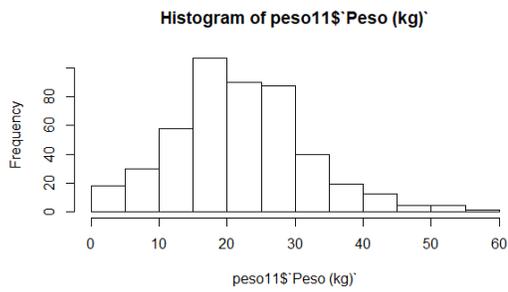
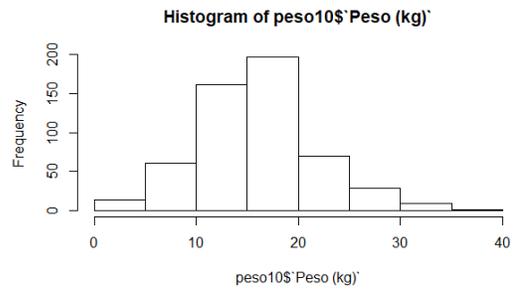
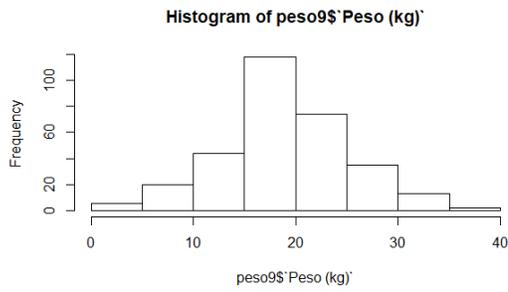
No.	Peso (kg)	Lote	No.	Peso (kg)	Lote
1			38		
2			39		
3			40		
4			41		
5			42		
6			43		
7			44		
8			45		
9			46		
10			47		
11			48		
12			49		
13			50		
14			51		
15			52		
16			53		
17			54		
18			55		

Fuente: elaboración propia.

Apéndice 2. Histogramas sin transformación BoxCox

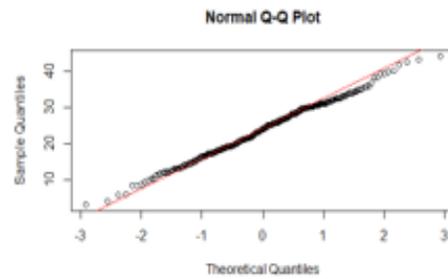
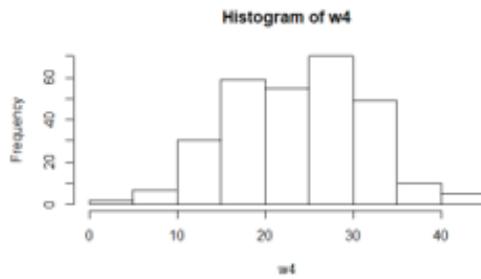
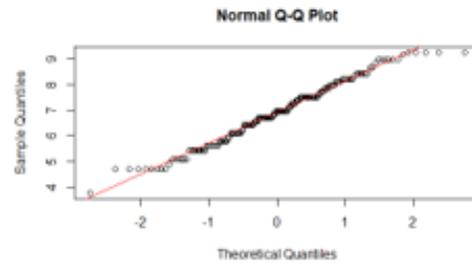
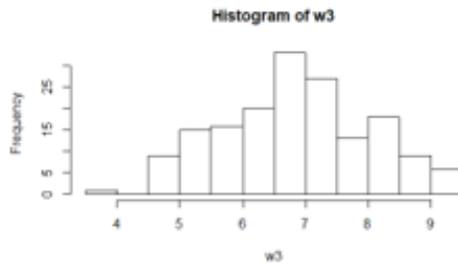
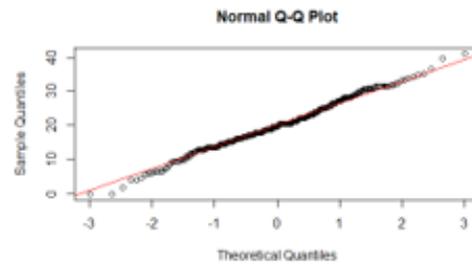
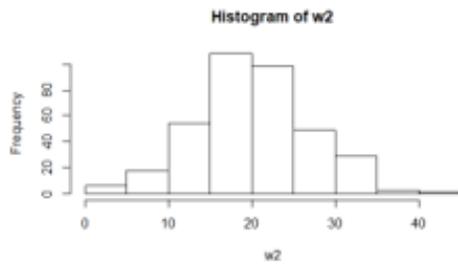
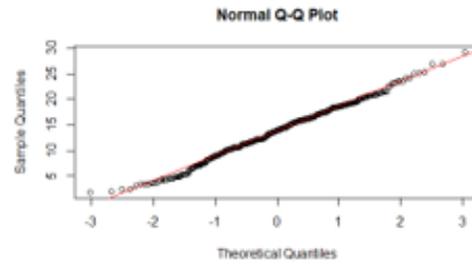
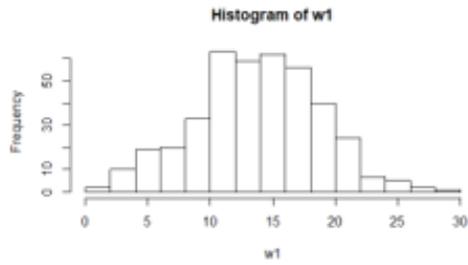


Continuación apéndice 2.

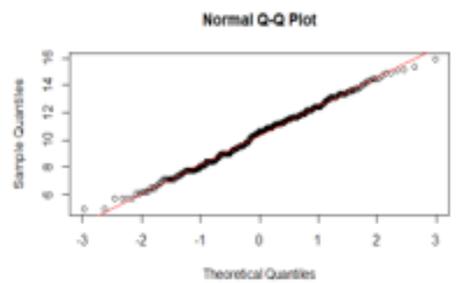
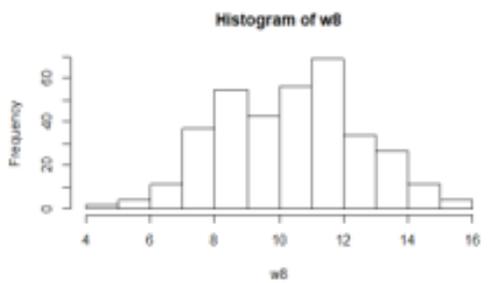
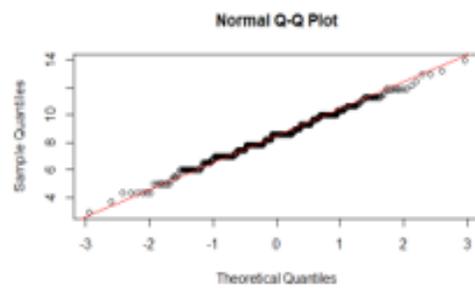
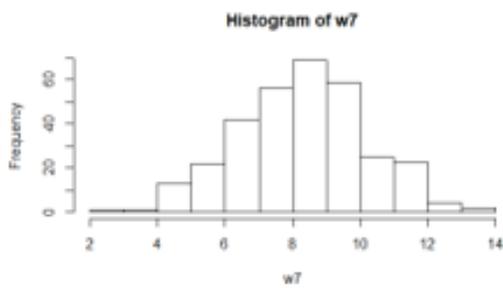
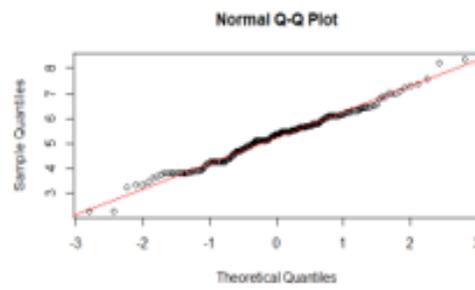
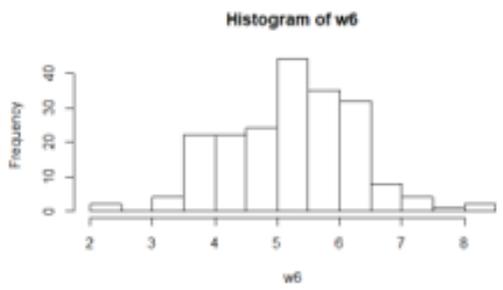
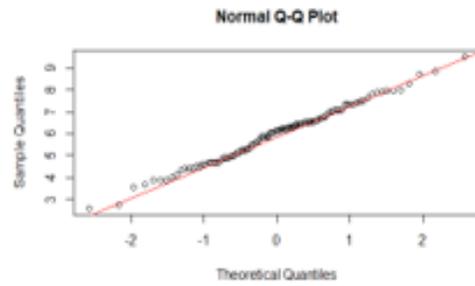
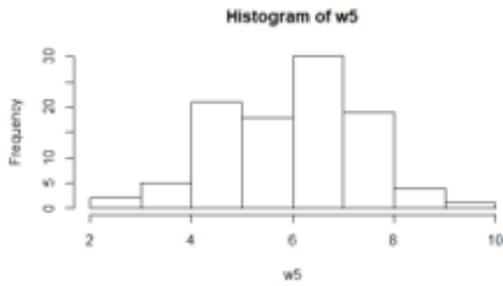


Fuente: elaboración propia, utilizando RStudio.

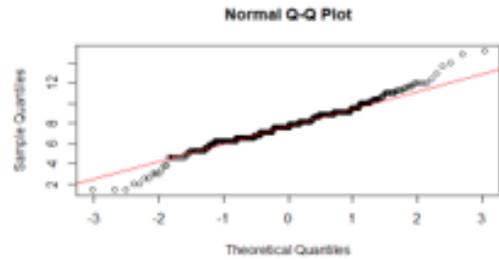
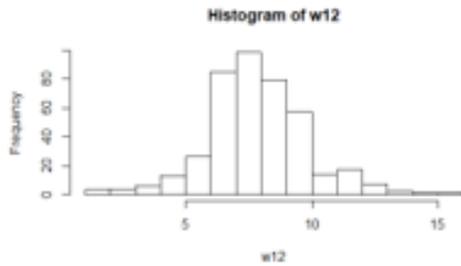
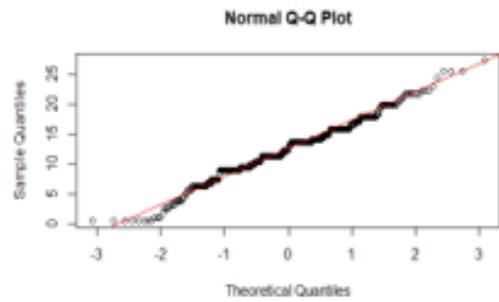
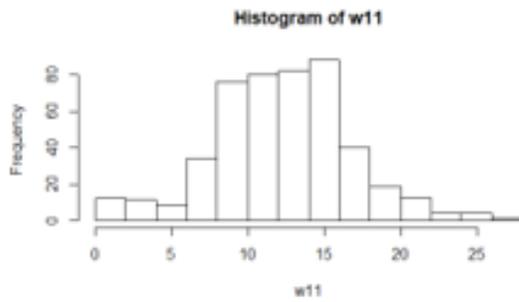
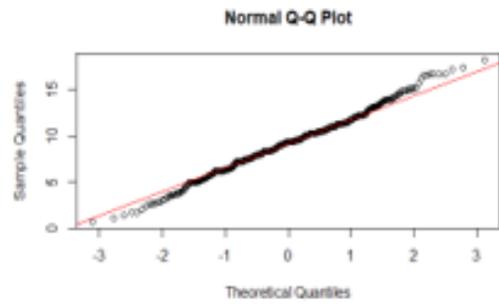
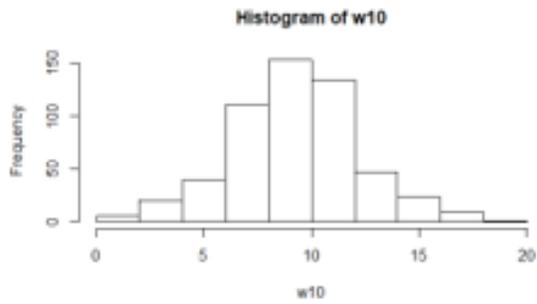
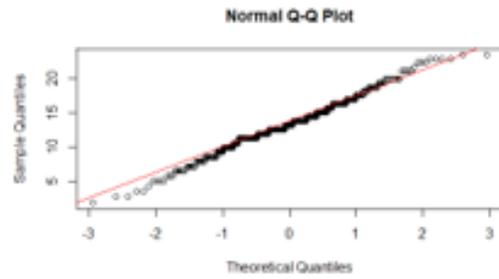
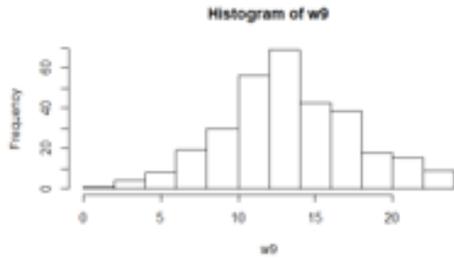
Apéndice 3. Histogramas y qqnorm después de BoxCox



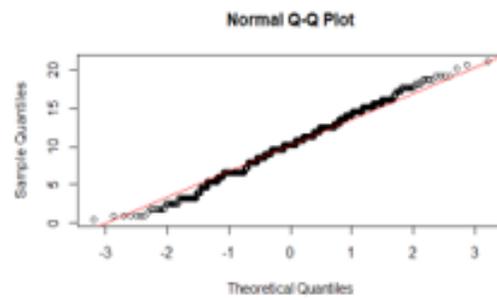
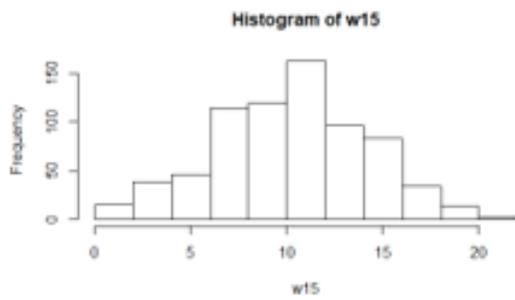
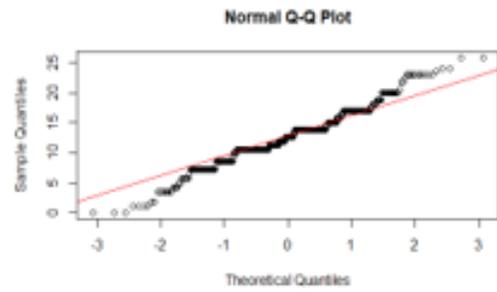
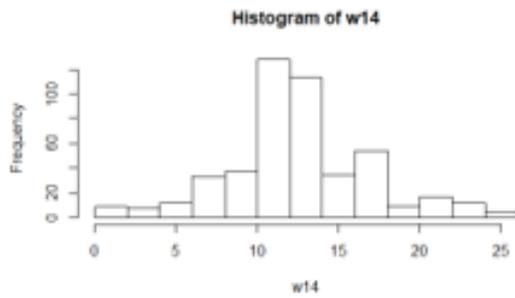
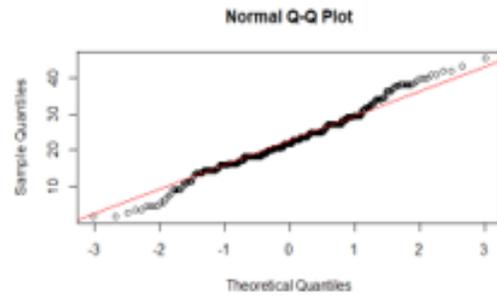
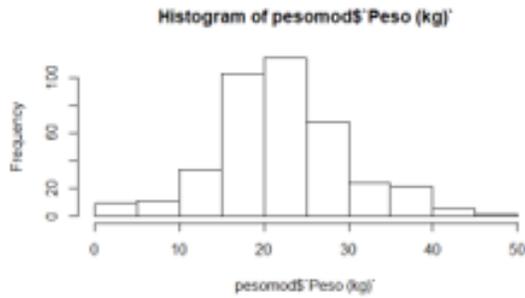
Continuación apéndice 3.



Continuación apéndice 3.



Continuación apéndice 3.



Fuente: elaboración propia, utilizando RStudio.

ANEXOS

Anexo 1. Distribución X²

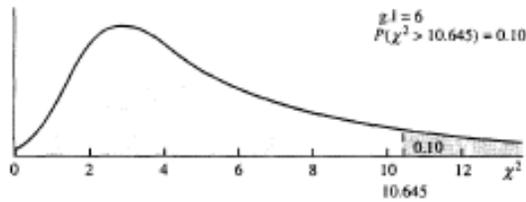


Tabla H Distribución chi-cuadrado

g.l.	$\chi^2_{0.999}$	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.980}$	$\chi^2_{0.970}$	$\chi^2_{0.960}$	$\chi^2_{0.950}$	$\chi^2_{0.940}$	$\chi^2_{0.930}$	$\chi^2_{0.920}$	$\chi^2_{0.910}$	$\chi^2_{0.900}$	$\chi^2_{0.850}$	$\chi^2_{0.800}$	$\chi^2_{0.750}$
1	0.000	0.000	0.001	0.004	0.016	0.148	0.455	1.074	1.642	2.706	3.841	5.024	5.412	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.713	1.386	2.408	3.219	4.605	5.991	7.378	7.824	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.424	2.366	3.665	4.642	6.251	7.815	9.348	9.837	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	2.195	3.357	4.878	5.989	7.779	9.488	11.143	11.668	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	3.000	4.351	6.064	7.289	9.236	11.070	12.833	13.388	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.828	5.348	7.231	8.558	10.645	12.592	14.449	15.033	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.671	6.346	8.383	9.803	12.017	14.067	16.013	16.622	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.527	7.344	9.524	11.030	13.362	15.507	17.535	18.168	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	6.393	8.343	10.656	12.242	14.684	16.919	19.023	19.679	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	7.267	9.342	11.781	13.442	15.987	18.307	20.483	21.161	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	8.148	10.341	12.899	14.631	17.275	19.675	21.920	22.618	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	9.034	11.340	14.011	15.812	18.549	21.026	23.337	24.054	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	9.926	12.340	15.119	16.985	19.812	22.362	24.736	25.472	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.821	13.339	16.222	18.151	21.064	23.685	26.119	26.873	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.721	14.339	17.322	19.311	22.307	24.996	27.488	28.259	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	12.624	15.338	18.418	20.465	23.542	26.296	28.845	29.633	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	13.531	16.338	19.511	21.615	24.769	27.587	30.191	30.995	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	14.440	17.338	20.601	22.760	25.989	28.869	31.526	32.346	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	15.352	18.338	21.689	23.900	27.204	30.144	32.852	33.687	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	16.266	19.337	22.775	25.038	28.412	31.410	34.170	35.020	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	17.182	20.337	23.858	26.171	29.615	32.671	35.479	36.343	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	18.101	21.337	24.939	27.301	30.813	33.924	36.781	37.659	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	19.021	22.337	26.018	28.429	32.007	35.172	38.076	38.968	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.943	23.337	27.096	29.553	33.196	36.415	39.364	40.270	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	20.867	24.337	28.172	30.675	34.382	37.652	40.646	41.566	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	21.792	25.336	29.246	31.795	35.563	38.885	41.923	42.856	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	22.719	26.336	30.319	32.912	36.741	40.113	43.194	44.140	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	23.647	27.336	31.391	34.027	37.916	41.337	44.461	45.419	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	24.577	28.336	32.461	35.139	39.087	42.557	45.722	46.693	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	25.508	29.336	33.530	36.250	40.256	43.773	46.979	47.962	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	34.872	39.335	44.165	47.269	51.805	55.758	59.342	60.436	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	44.313	49.335	54.723	58.164	63.167	67.505	71.420	72.613	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	53.809	59.335	65.227	68.972	74.397	79.082	83.298	84.580	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	63.346	69.334	75.689	79.715	85.527	90.531	95.023	96.388	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	72.915	79.334	86.120	90.405	96.578	101.879	106.629	108.069	112.329	116.321
90	59.195	61.754	65.647	69.126	73.291	82.511	89.334	96.524	101.054	107.565	113.145	118.136	119.648	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	92.129	99.334	106.906	111.667	118.498	124.342	129.561	131.142	135.807	140.169

Fuente: Webster A (2001). *Distribución X²* (p. 616).

Anexo 2. **Valores críticos de Dn para prueba de Kolmogorov-Smirnov**

Tamaño de muestra, n	Nivel de significancia para $D = \text{máximo } F_o - F_e $				
	0.20	0.15	0.10	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.510	0.565	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.368	0.410	0.490
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.450
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
más de 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Nota: Los valores de D dados en la tabla son valores críticos asociados con valores elegidos de n . Cualquier valor de D mayor o igual al valor de la tabla es significativo en el nivel de significancia indicado.

* Adaptado de F.J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit", *J. Am. Stat. Assoc.* 46:68-78, 1951. Con licencia del autor y de los editores.

Fuente: Levin y Rubin (2004). *Valores críticos de D para prueba de bondad de ajuste de Kolmogorov-Smirnov* (p. AT-23).

Anexo 3. Área bajo la curva de distribución normal

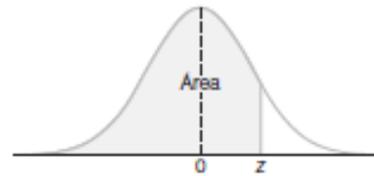


Tabla A.3 Áreas bajo la curva normal

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Continuación anexo 3.

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Fuente: Walpole et al. (2012) *Área bajo la curva de distribución normal* (p. 735).