

**Aplicación de técnicas de análisis de datos y programación para evaluar el efecto de la
disminución de la densidad de siembra en el cultivo de Brassicas en la empresa Cultivos
Casablanca**

Cristian Camilo Montaña Perez

Asesor

Andrea Ximena Robles

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Sociales Artes y Humanidades ECSAH
Especialización en Ciencia de Datos y Analítica
2025

Andrea Ximena Robles

Jurado

Jurado

Dedicatoria

A mi amada familia,

A mi esposa, por su amor incondicional, su paciencia infinita y por ser mi compañera constante en este camino lleno de desafíos y aprendizajes. Gracias por tu apoyo, por tus palabras de aliento en los momentos más difíciles y por estar siempre a mi lado, celebrando cada pequeño logro con entusiasmo y fe.

A mis hijos, quienes con su ternura, comprensión y alegría llenaron mis días de energía y esperanza. Gracias por su amor silencioso, por entender mis ausencias y por ser esa motivación inquebrantable que me impulsó a dar lo mejor de mí en cada etapa de este proceso.

Agradecimientos

Expreso mi más sincero agradecimiento a la empresa Cultivos Casablanca S.A.S. por haber facilitado el acceso a la información necesaria para el desarrollo de esta investigación, así como por su disposición para permitir la realización del ensayo en sus instalaciones. Su compromiso con la innovación y la mejora continua fue clave para la ejecución de este trabajo.

Agradezco también a los tutores y asesores del programa de Especialización en Ciencia de Datos y Analítica de la UNAD, en especial a la asesora Andrea Ximena Robles, por su orientación constante, sus aportes técnicos y su disposición para resolver inquietudes a lo largo del proceso.

Su acompañamiento fue fundamental para consolidar una propuesta metodológica sólida y coherente.

Finalmente, reconozco el respaldo institucional y académico recibido, el cual permitió vincular el conocimiento técnico con una problemática real del sector productivo, fortaleciendo el valor aplicado de este estudio.

Resumen

Este proyecto implementa un análisis estadístico apoyado en técnicas de ciencia de datos y programación para evaluar los resultados de un diseño de experimentos enfocado en el uso de diferentes densidades de siembra en el cultivo de flores Brassicas, desarrollado por la empresa Cultivos Casablanca S.A.S. El estudio contempla variables agronómicas clave como la longitud y el peso del tallo, el diámetro de la flor, las fechas de inicio y fin de la cosecha, la duración del periodo de cosecha, así como la relación entre el número de plantas cosechadas y el número de plantas sembradas. El objetivo es identificar patrones y efectos significativos que permitan optimizar las prácticas de siembra y mejorar la eficiencia productiva del cultivo.

Palabras clave: Diseño de experimentos, *Brassicas*, siembra

Abstract

This project implements a statistical analysis supported by data science and programming techniques to evaluate the results of an experimental design focused on the use of different planting densities in the cultivation of Brassica flowers, developed by Cultivos Casablanca S.A.S. The study considers key agronomic variables such as stem length and weight, flower diameter, harvest start and end dates, harvest period length, and the relationship between the number of plants harvested and the number of plants planted. The objective is to identify significant patterns and effects that allow for the optimization of planting practices and improved crop production efficiency.

Keywords: Experimental design, Brassicas, seedtime

Tabla de Contenido

Introducción	14
Justificación	15
Objetivos.....	16
Objetivo General.....	16
Objetivos Específicos	16
Marco de Referencia	17
Estado del Arte	17
Marco Contextual	18
Marco Teórico	20
Python	20
Bibliotecas	21
Pandas.....	21
SciPy.....	21
Statsmodels.....	22
NumPy.....	22
Seaborn.....	23
Matplotlib.....	23
Scikit-learn.....	24
Diseño de Experimentos	24
Prueba de Shapiro Wilk	25
Matriz de Correlación	26
Análisis de Varianza	27

Prueba de Kruskall - Wallis	28
Marco Conceptual.....	28
Kale Ornamental o Brassica Ornamental	29
Variedad Floral	29
Densidad de Siembra (Plantas/m ²).....	31
Cama de Producción	31
Aprovechamiento (%).....	31
Unidad Experimental	32
Variables Dependientes e Independientes	32
Python Como Herramienta de Análisis Estadístico.....	32
Marco Normativo	33
Metodología	35
Método.....	35
Tipo de Estudio.....	35
Recolección de Datos	35
Resultados	38
Tratamiento de Datos con Python	39
Tratamiento T0	42
% tallos aprovechados	42
Inicio de Cosecha.....	43
Fin de Cosecha.....	44
Spam de Cosecha	45
Diámetro de Cabeza.....	46

Diámetro de Tallo	47
Longitud de Tallo	48
Peso por Tallo	49
Tratamiento T1	50
% Tallos Aprovechados	50
Inicio de Cosecha.....	51
Fin de Cosecha.....	52
Spam de Cosecha	53
Diámetro de Cabeza.....	54
Diámetro de Tallo	55
Longitud de Tallo	56
Peso por Tallo	57
Tratamiento T2	58
% Tallos Aprovechados	58
Inicio de Cosecha.....	59
Fin de Cosecha.....	60
Spam de Cosecha	61
Diámetro de Cabeza.....	62
Diámetro de Tallo	63
Longitud de Tallo	64
Peso por Tallo	65
Matriz de Correlación de Kendall.....	68
Densidad de Siembra	69

	10
% Tallos Aprovechados	70
Inicio de Cosecha.....	70
Fin de Cosecha.....	71
Diámetro de Cabeza.....	71
Diámetro del Tallo	72
Longitud del Tallo	72
Peso por Tallo	72
Kruskal-Wallis	74
Post Hoc de Dunn	77
Conclusiones	84
Recomendaciones	87
Referencias.....	89

Lista de Figuras

Figura 1 <i>Variedad Crane Early White</i>	30
Figura 2 <i>Variedad Dream Red</i>	30
Figura 3 <i>Distribución de Plantas</i>	37
Figura 4 <i>Distribución de Unidades Experimentales en Campo</i>	37
Figura 5 <i>Librerías Usadas Para el Análisis en Python</i>	39
Figura 6 <i>Código en Python Para Importar y Leer la Base de Datos del Ensayo</i>	39
Figura 7 <i>Dataframe Generado con la Base de Datos del Ensayo</i>	39
Figura 8 <i>Código en Python Para Comprobar Supuesto de Normalidad y Homocedasticidad</i> ...	42
Figura 9 <i>Prueba de Normalidad % Aprovechamiento TO</i>	43
Figura 10 <i>Prueba de Normalidad Inicio de Cosecha Tratamiento TO</i>	44
Figura 11 <i>Prueba de Normalidad Fin de Cosecha Tratamiento TO</i>	45
Figura 12 <i>Prueba de Normalidad Spam de Cosecha Tratamiento TO</i>	46
Figura 13 <i>Prueba de Normalidad Diámetro de Cabeza Tratamiento TO</i>	47
Figura 14 <i>Prueba de Normalidad Diámetro de Tallo Tratamiento TO</i>	48
Figura 15 <i>Prueba de Normalidad Longitud de Tallo Tratamiento TO</i>	49
Figura 16 <i>Prueba de Normalidad Peso por Tallo Tratamiento TO</i>	50
Figura 17 <i>Prueba de Normalidad % Tallos de Aprovechamiento Tratamiento T1</i>	51
Figura 18 <i>Prueba de Normalidad Inicio de Cosecha Tratamiento T1</i>	52
Figura 19 <i>Prueba de Normalidad Fin de Cosecha Tratamiento T1</i>	53
Figura 20 <i>Prueba de Normalidad Spam de Cosecha Tratamiento T1</i>	54
Figura 21 <i>Prueba de Normalidad Diámetro de Cabeza Tratamiento T1</i>	55
Figura 22 <i>Prueba de Normalidad Diámetro de Tallo Tratamiento T1</i>	56

Figura 23 <i>Prueba de Normalidad Longitud de Tallo Tratamiento T1</i>	57
Figura 24 <i>Prueba de Normalidad Peso por Tallo Tratamiento T1</i>	58
Figura 25 <i>Prueba de Normalidad % Tallos Aprovechados Tratamiento T2</i>	59
Figura 26 <i>Prueba de Normalidad Inicio de Cosecha Tratamiento T2</i>	60
Figura 27 <i>Prueba de Normalidad Fin de Cosecha Tratamiento T2</i>	61
Figura 28 <i>Prueba de Normalidad Spam de Cosecha Tratamiento T2</i>	62
Figura 29 <i>Prueba de Normalidad Diámetro de Cabeza Tratamiento T2</i>	63
Figura 30 <i>Prueba de Normalidad Diámetro de Tallo Tratamiento T2</i>	64
Figura 31 <i>Prueba de Normalidad Longitud de Tallo Tratamiento T2</i>	65
Figura 32 <i>Prueba de Normalidad Peso por Tallo Tratamiento T2</i>	66
Figura 33 <i>Código en Python Para Construir Matriz de Correlación Método Kendall</i>	68
Figura 34 <i>Matriz de Correlación de Kendall</i>	69
Figura 35 <i>Código en Python Para Aplicar Prueba de Kruskal-Wallis</i>	75
Figura 36 <i>Resultado Prueba de Kruskal-Wallis</i>	76
Figura 37 <i>Código en Python Para Aplicar Prueba Post Hoc de Dunn</i>	78
Figura 38 <i>Resultado Prueba Post Hoc de Dunn Variable % Tallos Aprovechados</i>	78
Figura 39 <i>Resultado Prueba Post Hoc de Dunn Variable % Tallos Aprovechados</i>	79
Figura 40 <i>Resultado Prueba Post Hoc de Dunn Variable Diámetro de Cabezas</i>	79
Figura 41 <i>Resultado Prueba Post Hoc de Dunn Variable Diámetro de Tallo</i>	80
Figura 42 <i>Resultado Prueba Post Hoc de Dunn Variable Longitud de Tallos</i>	80
Figura 43 <i>Resultado Prueba Post Hoc de Dunn Variable Peso por Tallo</i>	81
Figura 44 <i>Código en Python Para Calcular Medias de Cada Variable por Tratamiento</i>	81
Figura 45 <i>Resultado Media por Variable por Tratamiento</i>	82

Lista de Tablas

Tabla 1 *Resumen Shapiro-Wilk Variables por Tratamiento* 67

Tabla 2 *Resumen Levene Variables por Tratamiento* 67

Introducción

El presente trabajo tiene como propósito evaluar el impacto de la disminución de la densidad de siembra en el cultivo de Brassicas ornamentales, empleando técnicas de análisis estadístico y programación. La investigación se desarrolló en la empresa Cultivos Casablanca S.A.S., un referente en la producción de flores diversificadas de exportación en la Sabana de Bogotá. Dada la creciente demanda por productos de alta calidad y la necesidad de optimizar los procesos productivos, se planteó un ensayo experimental que permitiera identificar cómo variaciones en la densidad de siembra inciden sobre parámetros agronómicos clave como el diámetro de la flor, el peso y la longitud del tallo, así como la eficiencia en el aprovechamiento de los tallos cosechables.

Para abordar esta problemática, se integraron herramientas de ciencia de datos y programación con un enfoque riguroso en el diseño experimental. La sistematización y análisis de los datos se realizó mediante Python, lo cual permitió una exploración robusta de las variables involucradas, la verificación de supuestos estadísticos y la aplicación de pruebas no paramétricas adaptadas a la naturaleza de los datos. El estudio se enmarca en un contexto técnico y económico relevante para el sector floricultor colombiano, contribuyendo con evidencia empírica al mejoramiento de las prácticas agronómicas y la toma de decisiones basadas en datos.

Justificación

La empresa Cultivos Casablanca se dedica a la producción y comercialización de flores frescas de exportación, destacándose como una finca especializada en la producción de especies diversificadas, es decir, distintas a la rosa y el clavel.

Dentro de su portafolio sobresale la Brassica, una variedad ornamental de apariencia llamativa y peculiar, cada vez más utilizada en arreglos florales y jardinería debido a su gran durabilidad. En Cultivos Casablanca, esta especie se cultiva en dos colores (púrpura y blanco) y representa un desafío constante cumplir con los exigentes parámetros de calidad requeridos por los clientes, entre ellos, la longitud del tallo, el peso del tallo y el diámetro de la flor.

Actualmente, la empresa destina aproximadamente 9,000 metros cuadrados al cultivo de Brassicas. Uno de los principales retos es mejorar la productividad, medida a través del aprovechamiento de tallos cosechados en relación con los tallos sembrados. En los últimos períodos, este indicador ha disminuido debido a la pérdida de tallos que no cumplen con los estándares de exportación, principalmente porque el diámetro de la flor no supera los 10 cm.

Se estima que reducir la densidad de siembra podría mejorar este aprovechamiento, ya que una menor competencia por recursos favorece la aireación, disminuye el riesgo de enfermedades, permite una mejor penetración de la luz y reduce el estrés fisiológico. Como resultado, las plantas pueden desarrollar una mejor estructura, acumular más biomasa, presentar menor incidencia de enfermedades y mejorar parámetros clave como el diámetro de la flor. Este estudio busca evaluar estos efectos con un enfoque basado en análisis de datos y técnicas estadísticas.

Objetivos

Objetivo General

Aplicar técnicas de análisis de datos y programación para evaluar el efecto de la disminución de la densidad de siembra en el cultivo de Brassicas en la empresa Cultivos Casablanca.

Objetivos Específicos

Analizar la relación entre la densidad de siembra y el diámetro de la flor en el cultivo de Brassicas mediante técnicas de análisis estadístico en Python.

Evaluar el impacto de la disminución de la densidad de siembra en el peso del tallo de las Brassicas utilizando herramientas de programación y análisis de datos.

Determinar el aprovechamiento de la siembra mediante la relación entre tallos cosechados y tallos sembrados, aplicando métodos estadísticos en Python.

Estimar el efecto de la densidad de siembra en el ciclo productivo de las Brassicas, considerando las variables de inicio, fin y duración de la cosecha mediante modelos estadísticos.

Utilizar métodos estadísticos en Python para determinar el impacto de la densidad de siembra en las variables de cosecha y su significancia.

Marco de Referencia

Estado del Arte

La densidad de siembra es uno de los factores agronómicos más relevantes en el manejo de cultivos, ya que influye directamente sobre el desarrollo morfológico, el rendimiento y la eficiencia del uso del espacio. En especies de interés comercial como las flores de corte o hortalizas ornamentales, ajustar la densidad puede marcar la diferencia entre un cultivo rentable y uno deficiente. Diversas investigaciones han explorado este aspecto en distintos tipos de plantas, aportando información valiosa sobre las respuestas fisiológicas y productivas a la variación en el número de plantas por metro cuadrado.

Kim et al. (2018) evaluaron el efecto de dos densidades de plantación sobre cuatro cultivos de repollo ornamental (*Brassica oleracea* var. *acephala*), concluyendo que una mayor densidad favorece la calidad visual del producto al generar rosetas más compactas y plantas más altas, aunque con menor biomasa total. Esta investigación constituye un referente fundamental para ensayos posteriores en brassicas ornamentales, como el presente, que busca optimizar la producción mediante ajustes de densidad.

En flores de corte, Abril Rodríguez et al. (2023) encontraron que aumentar la densidad de siembra en *Dianthus barbatus* hasta 54 plantas/m² permitió incrementar significativamente la productividad sin afectar negativamente la calidad, mostrando además un alto potencial de rentabilidad. Un hallazgo similar se dio en el cultivo de girasol ornamental, donde Chavarría Párraga y Vera Montenegro (2010) observaron que la mayor densidad (60 plantas/m²) generó el mayor número de flores por hectárea, aunque la mejor tasa de retorno económico se obtuvo con una densidad intermedia de 50 plantas/m². Ambos estudios evidencian cómo el equilibrio entre cantidad y calidad depende de la interacción entre densidad y genotipo.

Por su parte, Unigarro Muñoz et al. (2021) abordaron el tema desde una perspectiva fisiológica en café, relacionando la densidad de siembra con el índice de área foliar (IAF) y la interceptación de luz. Se concluyó que densidades altas permiten un cierre rápido del dosel vegetal y una mayor productividad por superficie, aunque podrían reducir la producción individual por planta. Este enfoque resalta el papel de la densidad como herramienta para maximizar la eficiencia fotosintética del cultivo.

En cultivos agrícolas como el frijol (*Phaseolus vulgaris* L.), Calero et al. (2018) demostraron que densidades intermedias (200 000 plantas/ha) lograron el mejor rendimiento en grano y variables morfofisiológicas, validando que la competencia entre plantas, cuando es bien manejada, puede beneficiar la productividad total.

A diferencia de los estudios anteriores, el presente ensayo experimental se enfoca en repollo ornamental bajo un rango más amplio de densidades y una medición de variables asociadas al valor comercial como peso, longitud, tamaño de cabeza y duración de cosecha, entre otras. Asimismo, incorpora herramientas de análisis estadístico moderno con Python, lo que permite una evaluación cuantitativa más precisa de los efectos de la densidad sobre el cultivo. Esta propuesta pretende no solo validar hallazgos previos en brassicas, sino también aportar evidencia nueva sobre la relación entre arquitectura vegetal y aprovechamiento comercial en este tipo de producto.

Marco Contextual

El presente estudio se desarrolla en el contexto del sector floricultor colombiano, específicamente en la empresa Cultivos Casablanca S.A.S., ubicada en el municipio de Chía, Cundinamarca. Esta zona forma parte de la Sabana de Bogotá, región que concentra alrededor del 70% de la producción nacional de flores, debido a sus condiciones agroclimáticas favorables

como altitud, temperatura constante, alta luminosidad y disponibilidad de agua (Asocolflores, 2023).

Colombia es el segundo exportador mundial de flores después de los Países Bajos, y el primero en claveles y crisantemos. El sector floricultor representa una parte significativa de la economía nacional, con más de 7.500 hectáreas cultivadas, generando más de 200.000 empleos directos e indirectos, y una participación cercana al 0,3% del PIB (DANE, 2022; ProColombia, 2023).

Según cifras de Asocolflores (2023), las exportaciones florales alcanzaron en 2022 un valor superior a USD 2.000 millones, con Estados Unidos como principal destino (más del 75%). Entre las variedades más representativas están el clavel, la rosa, el crisantemo, el alstroemeria y, en los últimos años, las flores diversificadas como las brassicas o kale, que han ganado importancia en mercados especializados por su valor ornamental y durabilidad poscosecha.

Cultivos Casablanca S.A.S. pertenece al Grupo GHT (Growers Hub Trading), uno de los conglomerados empresariales más reconocidos del sector., caracterizadas por su integración vertical, innovación genética y prácticas sostenibles certificadas (Grupo GHT, 2023). Estas empresas se han adaptado a las exigencias de los mercados internacionales a través de la diversificación de su portafolio floral.

El municipio de Chía, situado al norte de Bogotá, ha mostrado un desarrollo agroindustrial significativo, manteniendo la floricultura como una de sus actividades principales. Las condiciones climáticas temperaturas entre 10 °C y 18 °C, y altitud de más de 2.500 m s. n. permiten ciclos productivos constantes a lo largo del año (ICA, 2021).

En el caso particular de las brassicas o kales ornamentales, se trata de flores diversificadas pertenecientes a la familia Brassicaceae, cultivadas por su atractivo visual,

estructura simétrica y variedad de colores. Estas flores han sido incorporadas en la oferta de empresas como Cultivos Casablanca por su alta demanda en eventos, decoraciones y arreglos premium. Su producción requiere un manejo técnico cuidadoso, especialmente en etapas de siembra, desarrollo foliar, inducción floral y cosecha, siendo este estudio una contribución a la optimización del proceso productivo.

Este marco contextual permite entender no solo la relevancia del sector y su contribución a la economía nacional, sino también el entorno técnico y empresarial en el que se desarrollan los ensayos estadísticos y experimentales de la presente investigación.

Marco Teórico

Python

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general, que ha ganado gran popularidad en el ámbito científico y de análisis de datos. Su sintaxis clara y legible, junto con su amplia comunidad de usuarios, lo han convertido en una herramienta esencial para investigadores y analistas.

"Python se ha convertido en una opción predilecta para el análisis de datos debido a su simplicidad y la capacidad de integración con potentes bibliotecas científicas" (Martínez-Gómez, 2020, p. 45).

Además, su naturaleza de código abierto permite que la comunidad contribuya constantemente con mejoras, lo que ha derivado en el desarrollo de múltiples paquetes especializados en análisis estadístico y modelado.

"La flexibilidad de Python y la disponibilidad de bibliotecas como NumPy, Pandas y SciPy, han hecho posible su adopción masiva en proyectos científicos y de ingeniería de datos" (Oliphant, 2007).

Bibliotecas

Pandas. es una biblioteca desarrollada para proporcionar estructuras de datos rápidas, flexibles y expresivas diseñadas para trabajar con datos estructurados y de series temporales. Su principal estructura, el DataFrame, permite manipular grandes volúmenes de datos de manera eficiente.

"Pandas permite manipular, filtrar, agrupar y transformar datos con una sintaxis intuitiva, lo cual facilita el análisis exploratorio de datos" (McKinney, 2018, p. 4).

Pandas también proporciona funciones para leer y escribir archivos en múltiples formatos (CSV, Excel, SQL), facilitando la integración de fuentes de datos heterogéneas.

"La biblioteca Pandas es esencial para el trabajo diario de cualquier científico de datos que necesite limpiar, transformar y visualizar conjuntos de datos de manera eficiente" (VanderPlas, 2016, p. 92).

Adicional "Pandas ha transformado el panorama del análisis de datos, permitiendo realizar tareas complejas con pocas líneas de código y con una sintaxis accesible para usuarios no expertos" (Wes McKinney, 2012, p. 89).

SciPy. es una biblioteca que extiende las capacidades de NumPy y ofrece funciones optimizadas para álgebra lineal, estadística, integración numérica, y más. Su submódulo `scipy.stats` proporciona una amplia colección de herramientas para realizar pruebas estadísticas.

SciPy se ha consolidado como una de las bibliotecas científicas más importantes en Python, con funciones que cubren desde estadísticas básicas hasta modelos de optimización complejos" (Jones, Oliphant, & Peterson, 2001).

Esta biblioteca es ampliamente utilizada en ambientes académicos e industriales debido a su estabilidad y documentación exhaustiva.

"La biblioteca SciPy es parte fundamental del ecosistema científico de Python y su submódulo 'stats' permite realizar desde pruebas de hipótesis simples hasta análisis más avanzados" (Virtanen et al., 2020).

"El submódulo `scipy.stats` permite al usuario realizar contrastes de normalidad, correlaciones, análisis de varianza y ajustes de distribuciones con alta precisión y eficiencia computacional" (Millman & Aivazis, 2011).

Statsmodels. es una biblioteca orientada al modelado estadístico clásico, permitiendo realizar análisis de regresión lineal y no lineal, pruebas de hipótesis y modelos de series temporales.

"Statsmodels proporciona una interfaz clara para aplicar métodos estadísticos tradicionales dentro del entorno de programación Python, facilitando la replicación de análisis típicos en econometría y biostatística" (Seabold & Perktold, 2010).

Además, incluye herramientas para evaluar la calidad del ajuste de los modelos, como los residuos, el R^2 , los criterios AIC y BIC, lo que permite tomar decisiones fundamentadas en análisis empíricos.

"Es particularmente útil para realizar análisis estadísticos complejos con facilidad, replicando procedimientos que tradicionalmente se implementaban en software como R o Stata" (Davidson-Pilon, 2015).

"La documentación detallada y la similitud de su sintaxis con lenguajes estadísticos clásicos convierten a Statsmodels en una herramienta poderosa para usuarios que buscan rigor estadístico" (Brownlee, 2020).

NumPy. (Numerical Python) es una biblioteca fundamental para el análisis numérico y el cálculo científico en Python. Proporciona estructuras de datos eficientes, como los arreglos

multidimensionales (ndarray), que permiten realizar operaciones matemáticas y estadísticas de manera rápida y con bajo consumo de memoria. Su uso es común en aplicaciones de ciencia de datos, aprendizaje automático, procesamiento de señales y simulaciones científicas (Oliphant, 2006).

Una de las principales ventajas de NumPy es la posibilidad de vectorizar operaciones, lo cual elimina la necesidad de usar bucles explícitos y mejora significativamente el rendimiento computacional. Además, NumPy ofrece funciones avanzadas para álgebra lineal, transformadas de Fourier y generación de números aleatorios, lo que la convierte en una herramienta esencial en entornos de análisis y desarrollo científico (Van der Walt, Colbert, & Varoquaux, 2011).

Seaborn. es una biblioteca de visualización estadística basada en Matplotlib que permite generar gráficos informativos y atractivos con menos líneas de código. Está especialmente diseñada para trabajar de manera eficiente con estructuras de datos tipo DataFrame de Pandas.

"Seaborn proporciona una interfaz de alto nivel para dibujar gráficos estadísticos atractivos y bien informados" (Waskom, 2017, p. 2).

Además, Seaborn incluye funciones integradas para mostrar distribuciones, relaciones entre variables y análisis categóricos.

"La integración con Pandas y su enfoque en datos estadísticos convierte a Seaborn en una herramienta valiosa para el análisis exploratorio de datos" (VanderPlas, 2016, p. 101).

Matplotlib. es la biblioteca más antigua y flexible para la creación de gráficos en Python. Permite generar visualizaciones en 2D de alta calidad como gráficos de líneas, barras, dispersión, histogramas, entre otros.

"Matplotlib proporciona control total sobre cada aspecto de una figura, desde el tipo de línea hasta los textos, haciendo posible reproducir gráficamente casi cualquier tipo de visualización" (Hunter, 2007).

Gracias a su modularidad, Matplotlib puede ser usada tanto para visualizaciones rápidas como para gráficos complejos de publicación científica.

"Su arquitectura de bajo nivel permite que otras bibliotecas como Seaborn o Pandas plot se construyan sobre su base" (Tosi, 2020, p. 73).

Scikit-learn. es una biblioteca enfocada en el aprendizaje automático (machine learning), pero también incluye herramientas fundamentales para el análisis estadístico, como escalamiento de variables, selección de características y validación cruzada.

"Scikit-learn ofrece una amplia gama de algoritmos de aprendizaje supervisado y no supervisado, además de herramientas para la preprocesamiento de datos y evaluación de modelos" (Pedregosa et al., 2011).

Su integración con Pandas y NumPy la convierte en una opción eficiente y accesible para científicos de datos y analistas.

"La facilidad de uso y la coherencia en su interfaz la han convertido en una de las bibliotecas de machine learning más populares en Python" (Mueller & Massaron, 2016).

Diseño de Experimentos

El diseño de experimentos (DOE) es una metodología estadística que permite planificar, ejecutar y analizar experimentos de manera que se obtenga la máxima cantidad de información con el mínimo de recursos.

"El DOE permite identificar relaciones causa-efecto entre variables, optimizando la eficiencia en la recolección de datos" (Montgomery, 2012).

Se basa en principios como la aleatorización, la replicación y el control, lo cual garantiza la validez interna de los resultados.

"Los diseños factoriales permiten estudiar múltiples factores simultáneamente, revelando interacciones que serían invisibles con métodos univariados" (Kuehl, 2001).

Además, el DOE es fundamental para el análisis de procesos y mejora continua en entornos industriales.

"El diseño estadístico adecuado puede reducir significativamente el número de ensayos necesarios sin comprometer la calidad de las conclusiones" (Box, Hunter & Hunter, 2005).

Muchos métodos estadísticos clásicos, como ANOVA o regresión, requieren que los datos sigan una distribución normal. Verificar este supuesto es fundamental para evitar conclusiones erróneas.

"La normalidad de los residuos es uno de los supuestos clave en modelos estadísticos paramétricos" (Field, 2013).

El incumplimiento de este supuesto puede llevar a errores tipo I o II en pruebas de hipótesis, por lo que su verificación es paso crucial en el análisis de datos.

"Antes de aplicar pruebas paramétricas, se deben comprobar la normalidad de los datos y la homocedasticidad" (Hair et al., 2010).

También se ha documentado que las pruebas de normalidad deben complementarse con inspección gráfica (histogramas, Q-Q plots) para una evaluación robusta (Ghasemi & Zahediasl, 2012).

Prueba de Shapiro Wilk

La prueba de Shapiro-Wilk es uno de los métodos más utilizados para verificar la normalidad de una distribución. Es especialmente útil en muestras pequeñas.

"La prueba de Shapiro-Wilk es considerada una de las más potentes para detectar desviaciones de la normalidad" (Razali & Wah, 2011).

Se basa en la comparación entre los valores observados y los esperados bajo una distribución normal teórica. Si el valor p es menor que 0.05, se rechaza la hipótesis de normalidad.

"A pesar de su alta potencia, se recomienda acompañarla de análisis gráfico como histogramas o gráficos Q-Q" (Yap & Sim, 2011).

Prueba de Levene

La prueba de Levene es una técnica estadística utilizada para evaluar la igualdad de las varianzas entre dos o más grupos. Este supuesto es esencial en pruebas paramétricas como el ANOVA.

"La prueba de Levene es especialmente útil porque no asume normalidad en los datos, lo que la hace más robusta frente a distribuciones no normales" (Gastwirth, Gel, & Miao, 2009).

Su objetivo es comprobar la homogeneidad de la varianza (homocedasticidad) y se basa en analizar las desviaciones absolutas respecto a la mediana o la media de cada grupo.

"Cuando se detecta heterocedasticidad, el uso de pruebas no paramétricas o correcciones como Welch es más apropiado" (Tabachnick & Fidell, 2013).

Matriz de Correlación

Una matriz de correlación representa la relación entre variables numéricas, mostrando cómo varían juntas. Es una herramienta clave en el análisis exploratorio de datos.

"La matriz de correlación permite identificar patrones de dependencia y multicolinealidad entre variables predictoras" (Hair et al., 2010).

El coeficiente de correlación de Pearson se utiliza cuando las variables son continuas y normalmente distribuidas, mientras que métodos como Spearman y Kendall son más adecuados para datos ordinales o no normales.

"Visualizar la matriz de correlación mediante mapas de calor facilita la interpretación de grandes cantidades de variables interrelacionadas" (Yamane, 2011).

Cuando no se cumplen los supuestos de normalidad o se trata de datos ordinales, se recurre a coeficientes no paramétricos como Spearman y Kendall.

"La correlación de Spearman evalúa la relación monótona entre dos variables mediante el uso de rangos" (Zar, 2010).

Por su parte, el coeficiente de Kendall se basa en el conteo de concordancias y discordancias entre pares de observaciones, siendo más conservador en presencia de empates. "Kendall es más robusto cuando se tienen muchos valores empatados, aunque Spearman suele tener mayor potencia" (Conover, 1999).

Análisis de Varianza

El análisis de varianza (ANOVA) permite comparar las medias de tres o más grupos para identificar si existen diferencias estadísticamente significativas entre ellos.

"ANOVA es un procedimiento robusto basado en la partición de la varianza total en componentes atribuibles al tratamiento y al error" (Montgomery, 2017).

Es un método paramétrico que requiere el cumplimiento de supuestos como normalidad y homogeneidad de varianzas. Cuando se viola alguno, se debe considerar un ANOVA robusto o no paramétrico.

"La utilidad del ANOVA radica en su capacidad para evaluar múltiples comparaciones en una sola prueba, reduciendo el error tipo I" (Field, 2013).

Prueba de Kruskal - Wallis

La prueba de Kruskal-Wallis es una alternativa no paramétrica al ANOVA de un factor cuando los datos no cumplen los supuestos de normalidad o varianza homogénea.

"Kruskal-Wallis compara medianas entre grupos usando rangos, siendo adecuada para escalas ordinales o cuando la distribución de los datos es desconocida" (McKight & Najab, 2010).

Este test se basa en la suma de los rangos de cada grupo, y puede indicar si al menos uno de los grupos difiere significativamente del resto.

"Aunque no especifica cuáles grupos difieren, puede combinarse con pruebas post hoc para análisis más detallados" (Sheskin, 2004).

Prueba Post Hoc de Dunn

Cuando la prueba de Kruskal-Wallis detecta diferencias significativas, se puede aplicar una prueba post hoc de Dunn para comparaciones múltiples entre grupos.

"La prueba de Dunn permite realizar comparaciones pareadas ajustando el nivel de significancia para evitar errores tipo I" (Dunn, 1964).

El ajuste de Holm es una técnica de corrección más poderosa que Bonferroni, al aplicar un método secuencial que controla el error familiarizado (FWER). "El método de Holm proporciona un balance entre control de error y potencia estadística, siendo preferido en análisis post hoc no paramétricos" (Holm, 1979).

Marco Conceptual

En el desarrollo de esta investigación, centrada en la evaluación de variables productivas en el cultivo de Brassica se consideran conceptos clave que permiten contextualizar y

operacionalizar el diseño experimental, el análisis estadístico y la interpretación de resultados. A continuación, se definen los términos relevantes en el marco de este estudio.

Kale Ornamental o Brassica Ornamental

La kale ornamental es una variedad de Brassica oleracea cultivada con fines decorativos, especialmente valorada en floricultura por sus hojas rizadas y colores vistosos. Aunque pertenece a la misma especie que la col de consumo, se selecciona por atributos visuales más que por calidad alimenticia, y se comercializa como flor de corte en mercados de exportación. En Colombia, este cultivo ha ganado importancia por su resistencia, rusticidad y valor agregado en épocas frías.

Variedad Floral

En el contexto del ensayo, la variedad floral hace referencia al tipo genético específico de kale evaluado. Las variedades pueden diferenciarse por coloración, forma de la flor, vigor, ciclo de producción, y adaptabilidad agronómica. Estas diferencias influyen directamente en el comportamiento productivo y en las características de calidad evaluadas.

Figura 1

Variedad Crane Early White

**Figura 2**

Variedad Dream Red



Densidad de Siembra (Plantas/m²)

La densidad de siembra se refiere al número de plantas establecidas por metro cuadrado de cama o terreno cultivado. Es una variable independiente clave en el experimento, ya que influye directamente en la competencia por recursos (luz, nutrientes, agua) y, por ende, en el rendimiento y calidad del cultivo. Se evaluaron tres densidades diferentes para observar su efecto en las variables morfológicas y productivas de las plantas.

Cama de Producción

Una cama de producción es una unidad física delimitada del terreno, acondicionada para el establecimiento y manejo de cultivos florales. En este estudio, cada cama fue utilizada como unidad experimental, sembrada con una combinación específica de variedad y densidad, y su delimitación permitió controlar el manejo agronómico y registrar adecuadamente las variables de respuesta.

Aprovechamiento (%)

El aprovechamiento es el porcentaje de plantas sembradas que logran producir flores aptas para exportación. Se calcula como la proporción entre el número de tallos exportables y el total de plantas sembradas por cama. Este indicador permite evaluar la eficiencia productiva del cultivo bajo cada tratamiento.

Características morfológicas: longitud del tallo, peso del tallo, diámetro de la flor. Estas tres variables permiten medir la calidad física de la flor producida:

Longitud del tallo (cm): mide el largo del tallo desde la base hasta la inserción de la flor.

Peso del tallo (g): cuantifica la masa total de cada flor comercializable.

Diámetro de la cabeza de flor (cm): determina el tamaño visual de la flor, importante para su valoración comercial.

Estas características están asociadas a los estándares exigidos por los mercados internacionales y se ven influenciadas por la variedad y la densidad de siembra.

Inicio de cosecha, fin de cosecha y spam de cosecha

Inicio de cosecha: se define como el primer día en que se registra una flor exportable dentro de la cama evaluada.

Fin de cosecha: se refiere al último día en que se obtiene una flor que cumple los requisitos de exportación.

Spam de cosecha: se calcula como la diferencia en días entre el inicio y el fin de cosecha, lo que permite determinar la duración del periodo productivo de cada tratamiento. Un spam corto puede representar una cosecha más concentrada y eficiente.

Unidad Experimental

En este estudio, la unidad experimental corresponde a cada cama sembrada con una combinación específica de variedad floral y densidad de siembra. Esta delimitación permite aplicar tratamientos diferenciados y realizar mediciones independientes en función de las variables analizadas.

Variables Dependientes e Independientes

Las variables independientes del experimento fueron la variedad de kale y la densidad de siembra. Las variables dependientes, es decir, aquellas afectadas por los tratamientos aplicados, incluyen el aprovechamiento, longitud y peso del tallo, diámetro de la flor, inicio y fin de cosecha, y spam de cosecha.

Python Como Herramienta de Análisis Estadístico

Se utilizó el lenguaje de programación Python como soporte para el procesamiento de datos y análisis estadístico. Esta herramienta permitió aplicar métodos descriptivos y pruebas

estadísticas como verificación de normalidad, matriz de correlación, prueba de Kruskal-Wallis y análisis post hoc de Dunn, lo que asegura la reproducibilidad del análisis y la transparencia en el tratamiento de los datos.

Marco Normativo

El sector floricultor colombiano está regulado por una serie de normativas nacionales e internacionales que buscan asegurar la calidad de las flores, promover la sostenibilidad ambiental y garantizar el bienestar de los trabajadores. A nivel nacional, el Ministerio de Agricultura y Desarrollo Rural (MADR) establece políticas agrícolas clave que afectan a las fincas floricultoras. Entre las normativas más relevantes, se destaca la Ley 16 de 1990, que regula el sector agrícola y promueve el desarrollo de la agricultura en Colombia, buscando mejorar la competitividad de los productos agrícolas, incluidos los cultivos florales. Asimismo, el Decreto 2191 de 1997 establece disposiciones sobre la organización y funcionamiento de las entidades encargadas de la producción agrícola, incluyendo los lineamientos para la exportación de flores.

Además, el Ministerio de Agricultura y Desarrollo Rural (MADR) promueve la implementación de políticas públicas para la modernización y sostenibilidad del sector agrícola, con programas que buscan mejorar la infraestructura, la tecnología y la capacitación de los agricultores, lo cual es esencial para el sector floricultor (MADR, 2022).

El Instituto Colombiano Agropecuario (ICA), como autoridad fitosanitaria en Colombia, regula la sanidad vegetal de las flores que se producen para la exportación, garantizando que estas cumplan con los requisitos internacionales de calidad y salud. La Resolución 00718 de 2013 establece las condiciones fitosanitarias para la exportación de productos agrícolas, incluidas las flores, y las regulaciones sobre el control de plagas y enfermedades que podrían afectar la

salud de los cultivos. Asimismo, el Decreto 1843 de 2015 del ICA, establece las directrices para la implementación de buenas prácticas fitosanitarias en el sector agrícola.

A nivel internacional, las certificaciones de sostenibilidad juegan un papel crucial en la floricultura colombiana. La Rainforest Alliance establece normativas para asegurar que las fincas floricultoras cumplan con estándares ambientales y sociales rigurosos, promoviendo la protección de los ecosistemas y la biodiversidad (Rainforest Alliance, 2020). De manera similar, la certificación Flor Verde regula las prácticas agrícolas responsables en el sector, abogando por un manejo adecuado de los recursos naturales, la reducción de la contaminación y el respeto por los derechos de los trabajadores (Flor Verde, 2021).

Por otro lado, el Código Sustantivo del Trabajo de Colombia regula los derechos laborales de los empleados del sector floricultor, asegurando que se cumplan los estándares de salud, seguridad y bienestar en el lugar de trabajo. Las normativas laborales son complementadas por las regulaciones de salud y seguridad ocupacional, que buscan proteger a los trabajadores expuestos a riesgos químicos y físicos en las fincas floricultoras (Ministerio de Trabajo, 2021)

Metodología

Método

La presente investigación se desarrolló bajo un enfoque cuantitativo, debido a que se utilizaron mediciones numéricas y herramientas estadísticas para evaluar el comportamiento de diferentes variables agronómicas en función de la densidad de siembra en el cultivo de flores Brassicas. Este enfoque permitió analizar de forma objetiva los efectos de los tratamientos aplicados, identificar tendencias y establecer relaciones entre variables.

Tipo de Estudio

Este estudio corresponde a una investigación de tipo experimental, ya que se implementó un diseño de experimentos en el que se manipularon deliberadamente los niveles de densidad de siembra para observar su efecto sobre variables de interés como la longitud del tallo, el peso del tallo, el diámetro de la flor, las fechas de inicio y fin de la cosecha, la duración del ciclo de recolección y la relación entre plantas cosechadas y sembradas.

Se definieron tres tratamientos para evaluar: T0 - 57 plantas por metro cuadrado bajo invernadero, T1 - 46 plantas por metro cuadrado bajo invernadero, T2-31 plantas por metro cuadrado bajo invernadero

El diseño experimental permite establecer una relación causa-efecto entre los tratamientos aplicados y las respuestas observadas, contribuyendo a la toma de decisiones técnicas basadas en evidencia dentro del proceso productivo.

Recolección de Datos

La recolección de datos se llevó a cabo en los cultivos de la empresa Cultivos Casablanca S.A.S., ubicados en la zona rural del municipio de Chía, Cundinamarca (coordenadas: 4.9061° N, 74.0691° O)], durante el ciclo de producción de Octubre a Enero de 2025.

En el cultivo de Brassicas en Cultivos Casablanca S.A.S., las siembras se realizan en unidades de producción conocidas como camas, que corresponden a espacios físicos delimitados dentro de los invernaderos. Cada cama tiene unas dimensiones de 1,3 metros de ancho por 30 metros de largo.

Durante la siembra y a lo largo del ciclo de vida del cultivo, se utiliza una malla plástica con cuadrícula de 10 cm x 10 cm. Esta malla cumple una doble función: primero, actúa como guía o regleta en el momento de la siembra, facilitando una distribución uniforme de las plantas; y segundo, durante el desarrollo del cultivo, sirve como sistema de tutorado, brindando soporte a los tallos para mantenerlos erguidos y favorecer su crecimiento recto, lo cual es esencial para la calidad del producto final.

En la siguiente figura se muestra la distribución de plantas sobre la malla plástica según el tratamiento aplicado. La malla cuenta con 13 espacios a lo ancho de la cama y 300 líneas a lo largo, formando una cuadrícula regular de 10 cm x 10 cm.

En el tratamiento T0, se siembran 11 plantas por línea, dejando un espacio vacío a cada lado de la cama para mantener simetría.

En el tratamiento T1, se distribuyen 9 plantas por línea, dejando 4 espacios vacíos (1 cada 3 plantas).

En el tratamiento T2, se siembran 6 plantas por línea, distribuidas de forma que haya 7 espacios vacíos, incluyendo un espacio intermedio entre planta y planta, lo que proporciona mayor separación y ventilación.

Esta configuración permite evaluar el efecto de diferentes densidades de siembra sobre el desarrollo del cultivo, manteniendo la uniformidad y precisión en la distribución de las plantas gracias a la estructura de la malla.

Figura 3*Distribución de Plantas*

	1	2	3	4	5	6	7	8	9	10	11	12	13
T0		●	●	●	●	●	●	●	●	●	●	●	
T1		●	●	●		●	●	●		●	●	●	
T2		●		●		●		●		●		●	

Nota. Para cada tratamiento se definieron 5 parcelas o unidades experimentales, como sigue a continuación.

Figura 4*Distribución de Unidades Experimentales en Campo*

T0	T1	T2
1 m (borde)	1 m (borde)	1 m (borde)
Parcela 1	Parcela 1	Parcela 1
Parcela 2	Parcela 2	Parcela 2
Parcela 3	Parcela 3	Parcela 3
Parcela 4	Parcela 4	Parcela 4
Parcela 5	Parcela 5	Parcela 5
1 m (borde)	1 m (borde)	1 m (borde)

En cada unidad experimental se registraron las siguientes variables: longitud del tallo

(cm) 40 muestras por cada parcela, peso del tallo (g) 20 muestras por cada parcela, diámetro de la flor (cm) 40 muestras por cada parcela, fecha de inicio de cosecha 5 muestras por tratamiento,

fecha de finalización de la cosecha 5 muestras por tratamiento, duración de la cosecha (días) 5 muestras por tratamiento, número de plantas sembradas por cada parcela, número de plantas cosechadas por cada parcela

La información fue recolectada de forma manual mediante formatos de campo previamente diseñados. Posteriormente, los datos fueron sistematizados utilizando hojas de cálculo y analizados mediante herramientas estadísticas con apoyo de lenguajes de programación como Python.

Resultados

Tratamiento de Datos con Python

Los datos recolectados durante el estudio son analizados a través de técnicas estadísticas usando programación y lenguajes como Python y R.

Se utilizan las siguientes librerías para el análisis de los datos

Figura 5

Librerías Usadas Para el Análisis en Python

```
import pandas as pd
import scipy.stats as stats
import statsmodels.api as sm
from statsmodels.formula.api import ols
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import scikit_posthocs as sp
```

Nota. Se importa la base de datos a través de la función `pd.read`, creando un dataframe denominado `df`.

Figura 6

Código en Python Para Importar y Leer la Base de Datos del Ensayo

```
df = pd.read_excel('/content/Datos Ensayo Kale.xlsx', sheet_name='Productividad')
df1 = df[['Variedad', 'Tratamiento', 'Densidad de siembra', '% tallos aprovechados', 'Inicio', 'Fin', 'Spam', 'DiametroCabeza', 'DiametroTallo', 'LongitudTallo', 'Peso por tallo']]
```

La base de datos que resume los resultados de los tratamientos se resume así:

Figura 7

Dataframe Generado con la Base de Datos del Ensayo

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 30 entries, 0 to 29
```

```
Data columns (total 25 columns):
```

#	Column	Non-Null Count	Dtype
0	Variedad	30 non-null	object
1	Tratamiento	30 non-null	object
2	Densidad de siembra	30 non-null	int64
3	Parcela	30 non-null	int64
4	plantas iniciales	30 non-null	int64
5	tallos aprovechados	30 non-null	int64
6	tallos NO aprovechados	30 non-null	int64
7	Cabeza pequeña	26 non-null	float64
8	Ramificado	21 non-null	float64
9	Daño mecánico	16 non-null	float64
10	Tallo corto	26 non-null	float64
11	Entorchado	25 non-null	float64
12	Tallo delgado	30 non-null	int64
13	Tallo torcido	25 non-null	float64
14	Otros	26 non-null	float64
15	Erradicaciones fitosanitarias	0 non-null	float64
16	% tallos aprovechados	30 non-null	float64
17	% de tallos NO aprovechados	30 non-null	float64
18	Inicio	30 non-null	int64
19	Fin	30 non-null	int64
20	Spam	30 non-null	int64
21	LongitudTallo	30 non-null	float64
22	DiametroTallo	30 non-null	float64
23	DiametroCabeza	30 non-null	float64
24	Peso por tallo	30 non-null	float64

```
dtypes: float64(14), int64(9), object(2)
```

```
memory usage: 6.0+ KB
```

Se define como variable independiente del ensayo el tratamiento densidad de siembra: T0 (57), T1 (46) , T2 (31) y como variables respuesta las variables dependientes: % tallos aprovechados, Inicio de Cosecha, Fin de Cosecha, Spam de cosecha, Longitud del tallo, Diámetro del tallo, Diámetro de cabeza y Peso por tallo.

Iniciamos el análisis comprobando supuesto de normalidad en cada uno de los grupos y variables definidas:

Se genera el siguiente código en Python para analizar cada variable en cada tratamiento, aplicando la prueba de Shapiro Wilk comprobando normalidad, prueba de Levene comprobando homocedasticidad y generando gráficos como histogramas y Q-Q plot.

Figura 8

Código en Python Para Comprobar Supuesto de Normalidad y Homocedasticidad

```
# Recorremos cada tratamiento
for tratamiento in tratamientos:
    print(f"\n\n--- Tratamiento: {columna_tratamiento} = {tratamiento} ---")

    # Filtrar solo los datos de este tratamiento
    df_trat = df1[df1[columna_tratamiento] == tratamiento]

    for columna in columnas_numericas:
        print(f"\nVariable: {columna}")

        datos = df_trat[columna].dropna()

        # 1. Prueba de Shapiro-Wilk
        stat, p = stats.shapiro(datos)
        print(f" Shapiro-Wilk: estadístico = {stat:.4f}, p-valor = {p:.4f}")
        if p > 0.05:
            print(" Distribución normal (no se rechaza H0)")
        else:
            print(" No es distribución normal (se rechaza H0)")

        # 2. Prueba de Levene (para homocedasticidad entre tratamientos)
        stat_levene, p_levene = stats.levene(*[df1[df1[columna_tratamiento] == t][columna].dropna() for t in tratamientos])
        cumple_homocedasticidad = "Sí" if p_levene > 0.05 else "No"
        resultados_levene[tratamiento][columna] = f"{cumple_homocedasticidad} (p={p_levene:.4f})"
        print(f" Prueba de Levene: estadístico = {stat_levene:.4f}, p-valor = {p_levene:.4f}")
        print(f" Cumple con homocedasticidad: {cumple_homocedasticidad}")

        # 3. Histograma con KDE
        plt.figure(figsize=(6, 4))
        sns.histplot(datos, kde=True, bins=10, color='skyblue')
        plt.title(f"Histograma - {columna} \nTratamiento: {tratamiento}")
        plt.xlabel(columna)
        plt.ylabel("Frecuencia")
        plt.grid(True)
        plt.tight_layout()
        plt.show()

        # 3. Q-Q Plot
        plt.figure(figsize=(6, 4))
        stats.probplot(datos, dist="norm", plot=plt)
        plt.title(f"Q-Q Plot - {columna} \nTratamiento: {tratamiento}")
        plt.grid(True)
        plt.tight_layout()
        plt.show()
```

Tratamiento T0

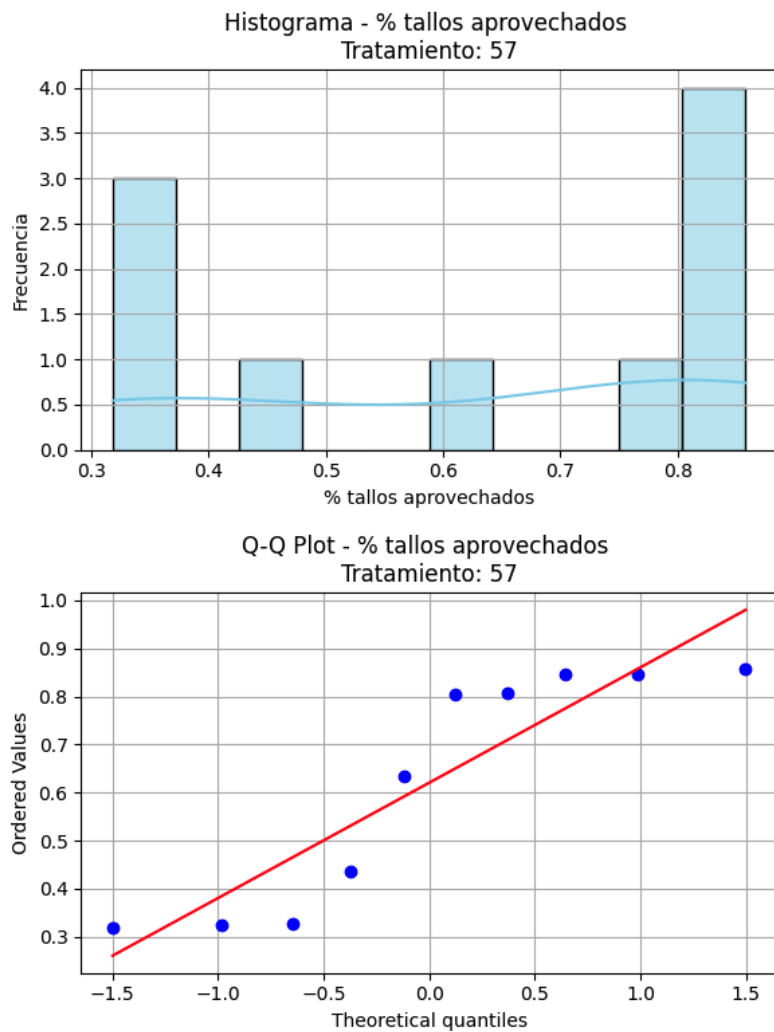
% tallos aprovechados

El resultado de la aplicación de código para la variable % de tallos aprovechados en el tratamiento T0. No cumple normalidad, no cumple homocedasticidad.

Figura 9

Prueba de Normalidad % Aprovechamiento T0

Variable: % tallos aprovechados
 Shapiro-Wilk: estadístico = 0.7916, p-valor = 0.0115
 No es distribución normal (se rechaza H0)
 Prueba de Levene: estadístico = 8.8961, p-valor = 0.0011
 Cumple con homocedasticidad: No

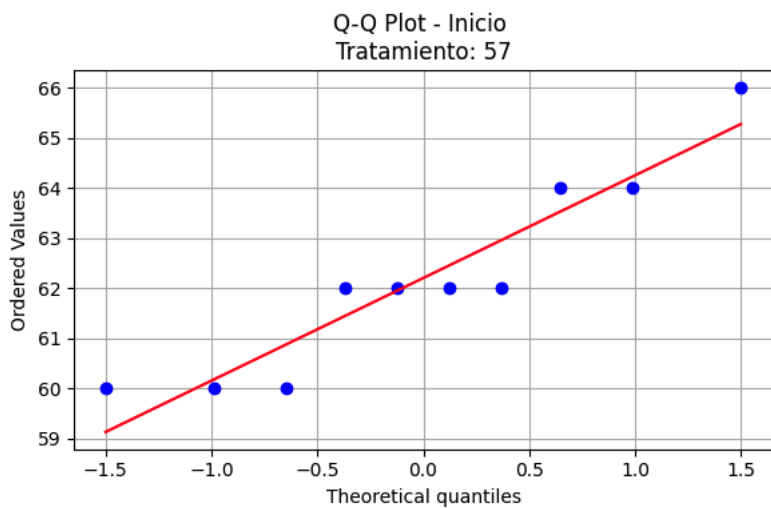
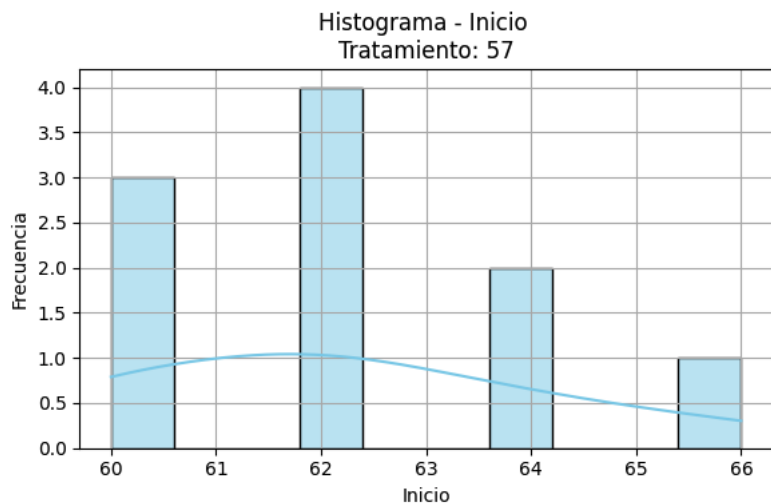


Inicio de Cosecha

El resultado de la aplicación de código para la variable inicio de cosecha en el tratamiento T0. Cumple normalidad, Cumple homocedasticidad.

Figura 10*Prueba de Normalidad Inicio de Cosecha Tratamiento T0*

Variable: Inicio
 Shapiro-Wilk: estadístico = 0.8858, p-valor = 0.1520
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 0.7532, p-valor = 0.4805
 Cumple con homocedasticidad: Sí

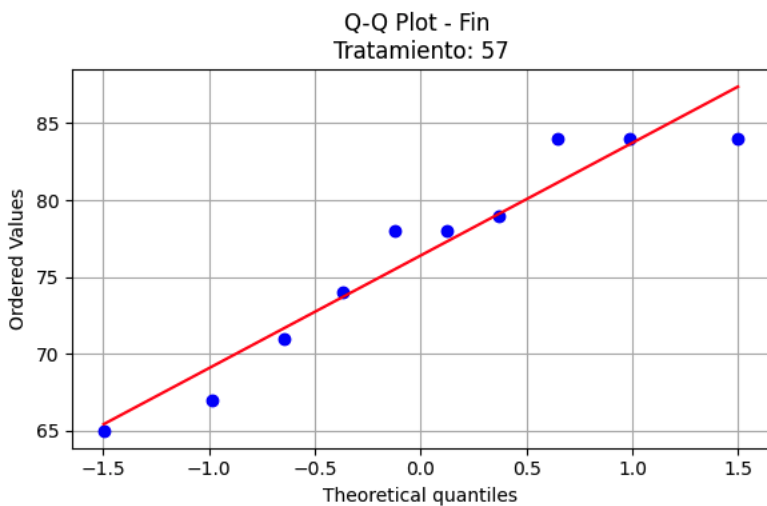
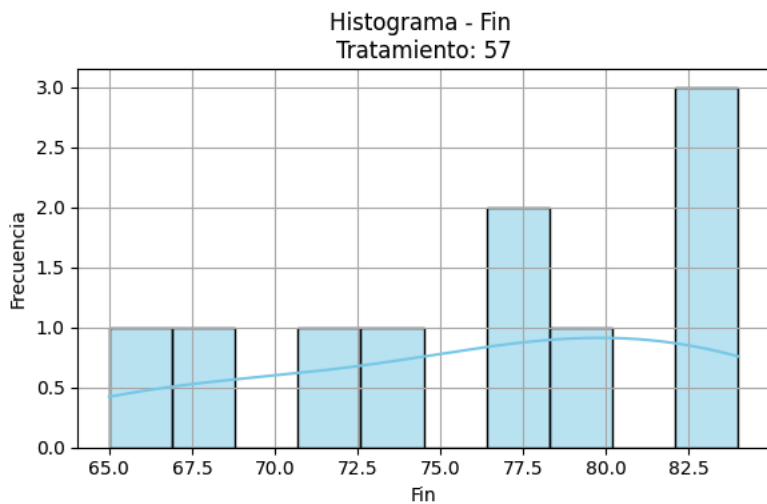
*Fin de Cosecha*

El resultado de la aplicación de código para la variable Fin de cosecha en el tratamiento

T0. Cumple normalidad, Cumple homocedasticidad.

Figura 11*Prueba de Normalidad Fin de Cosecha Tratamiento T0*

Variable: Fin
 Shapiro-Wilk: estadístico = 0.9027, p-valor = 0.2346
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 2.4334, p-valor = 0.1067
 Cumple con homocedasticidad: Sí

*Spam de Cosecha*

El resultado de la aplicación de código para la variable Spam de cosecha en el tratamiento T0. Cumple normalidad, Cumple homocedasticidad.

Figura 12*Prueba de Normalidad Spam de Cosecha Tratamiento T0*

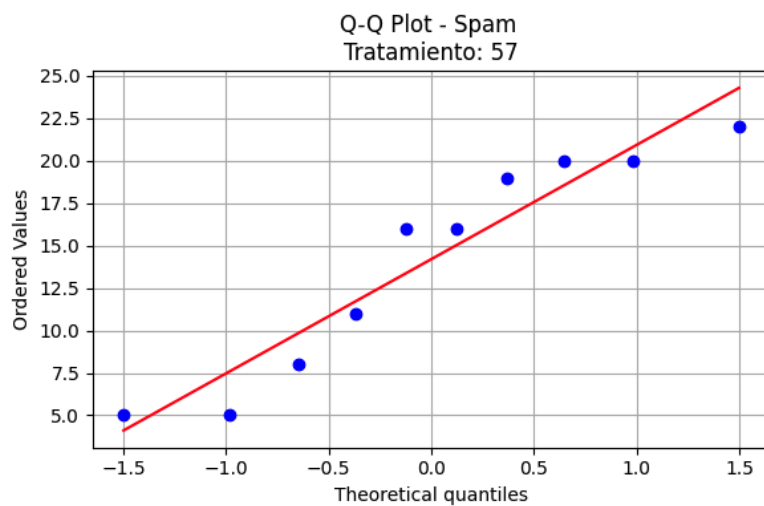
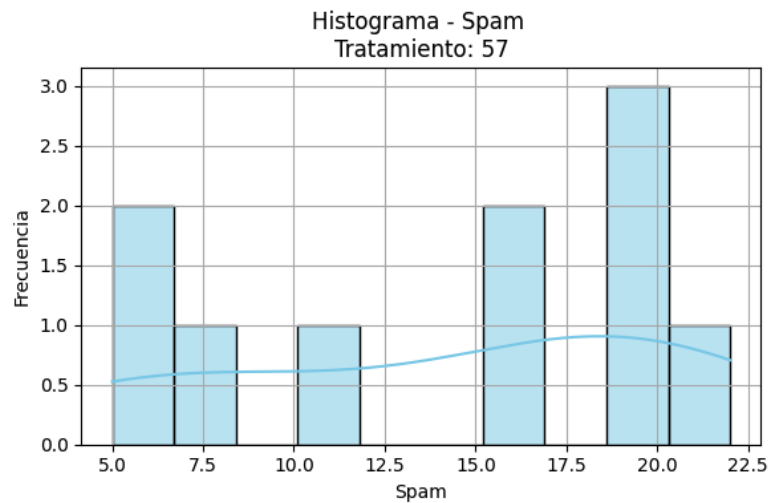
Variable: Spam

Shapiro-Wilk: estadístico = 0.8860, p-valor = 0.1528

Distribución normal (no se rechaza H0)

Prueba de Levene: estadístico = 0.7937, p-valor = 0.4624

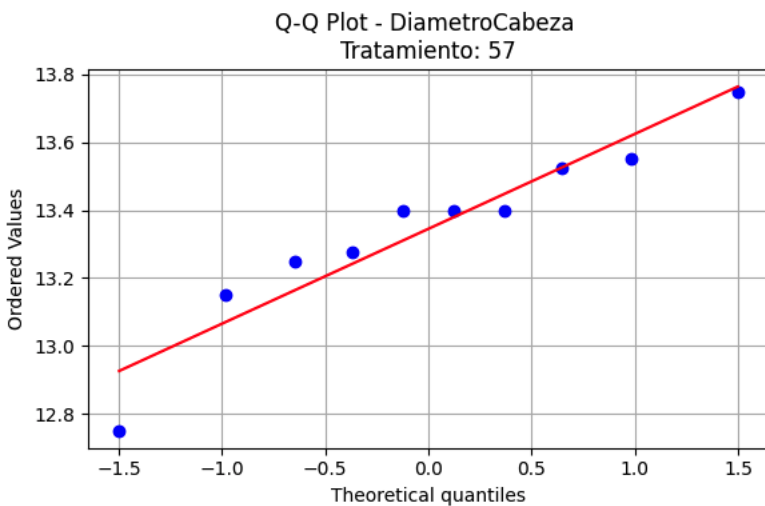
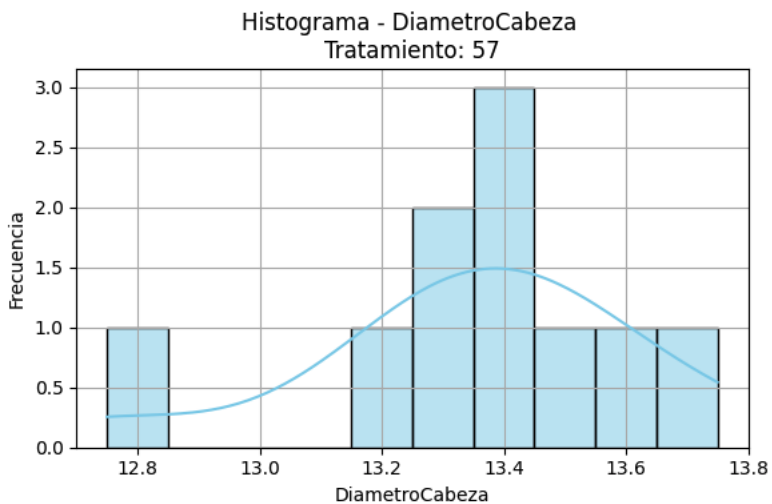
Cumple con homocedasticidad: Sí

*Diámetro de Cabeza*

El resultado de la aplicación de código para la variable Diámetro de cabeza en el tratamiento T0. Cumple normalidad, no cumple homocedasticidad.

Figura 13**Prueba de Normalidad Diámetro de Cabeza Tratamiento TO**

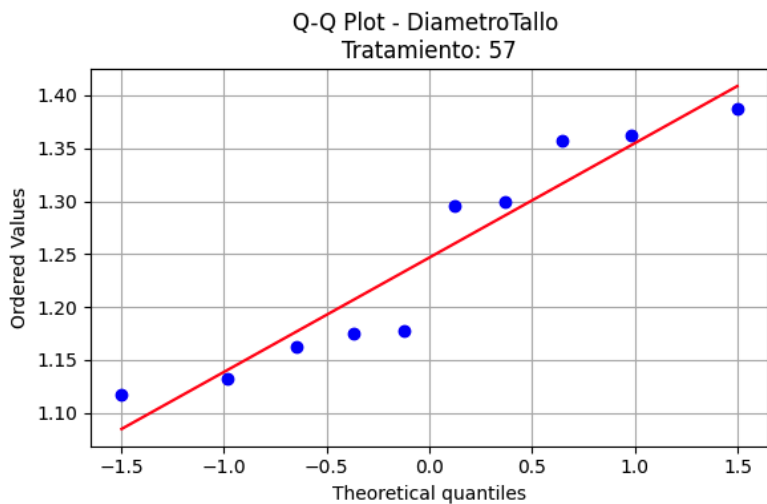
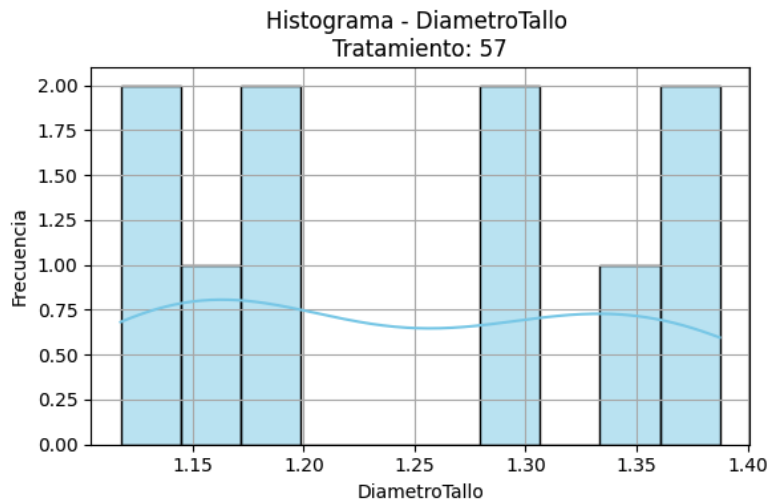
Variable: DiametroCabeza
 Shapiro-Wilk: estadístico = 0.9300, p-valor = 0.4482
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 18.1369, p-valor = 0.0000
 Cumple con homocedasticidad: No

***Diámetro de Tallo***

El resultado de la aplicación de código para la variable diámetro de tallo en el tratamiento T0. Cumple normalidad, Cumple homocedasticidad.

Figura 14*Prueba de Normalidad Diámetro de Tallo Tratamiento TO*

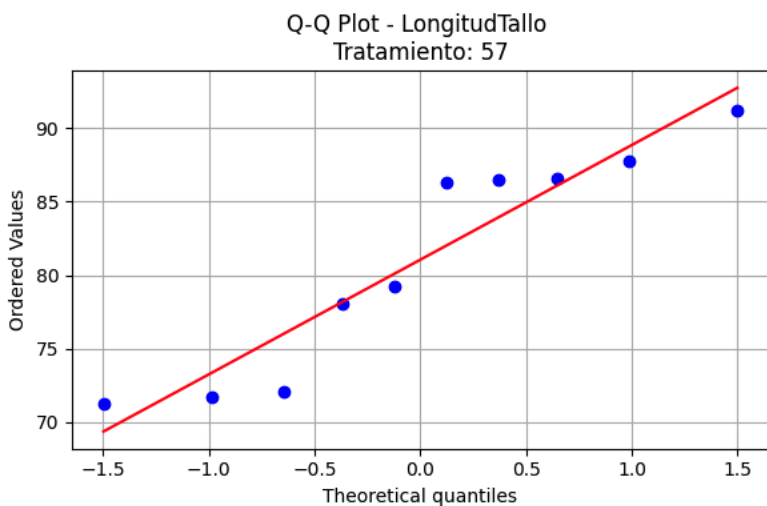
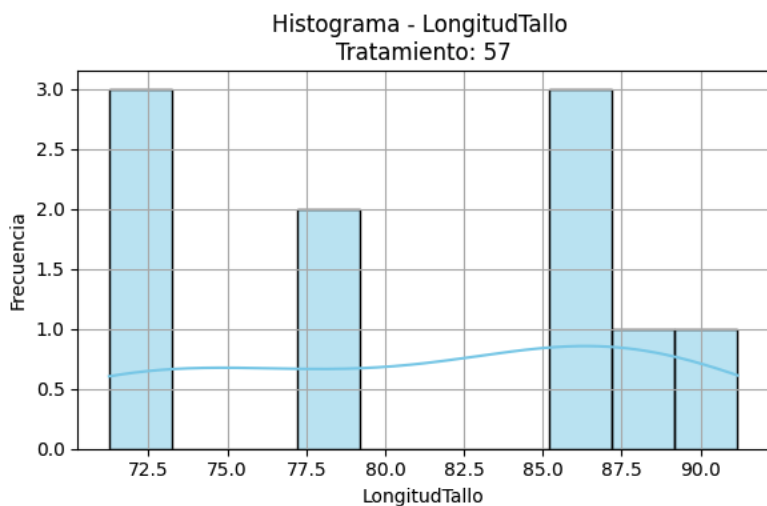
Variable: DiametroTallo
 Shapiro-Wilk: estadístico = 0.8776, p-valor = 0.1224
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 1.2655, p-valor = 0.2983
 Cumple con homocedasticidad: Sí

***Longitud de Tallo***

El resultado de la aplicación de código para la variable longitud de tallo en el tratamiento T0. Cumple normalidad, no cumple homocedasticidad.

Figura 15*Prueba de Normalidad Longitud de Tallo Tratamiento T0*

Variable: LongitudTallo
 Shapiro-Wilk: estadístico = 0.8737, p-valor = 0.1103
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 12.0861, p-valor = 0.0002
 Cumple con homocedasticidad: No

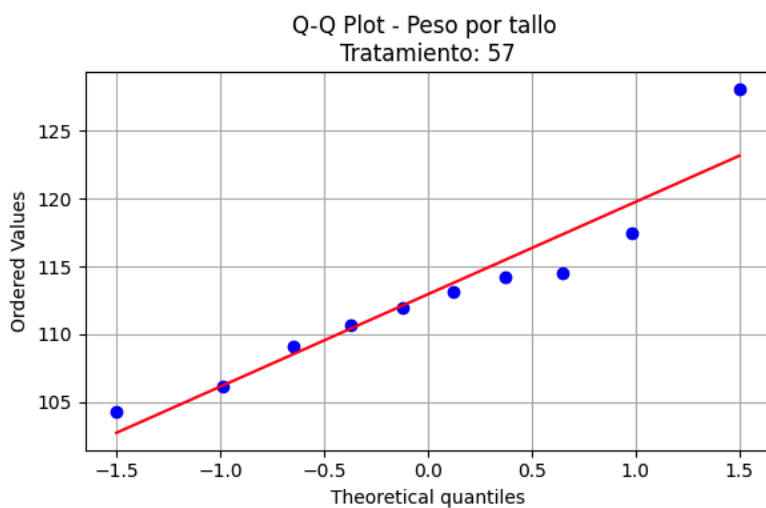
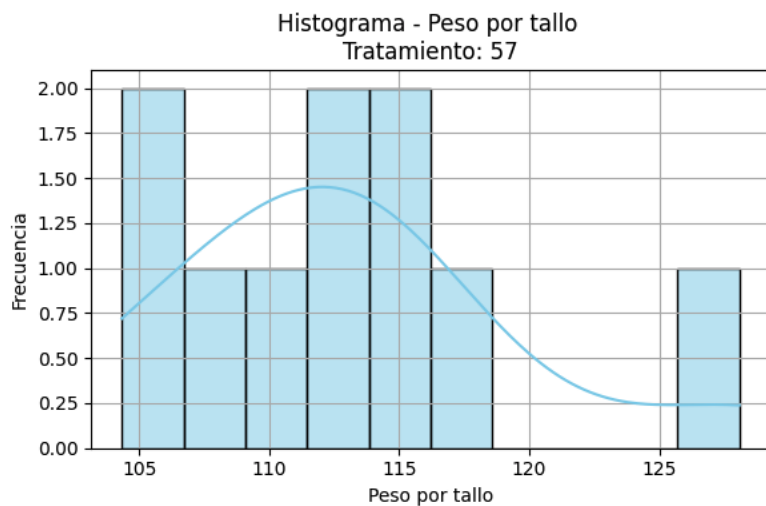
***Peso por Tallo***

El resultado de la aplicación de código para la variable peso por tallo en el tratamiento

T0. Cumple normalidad, no cumple homocedasticidad.

Figura 16*Prueba de Normalidad Peso por Tallo Tratamiento TO*

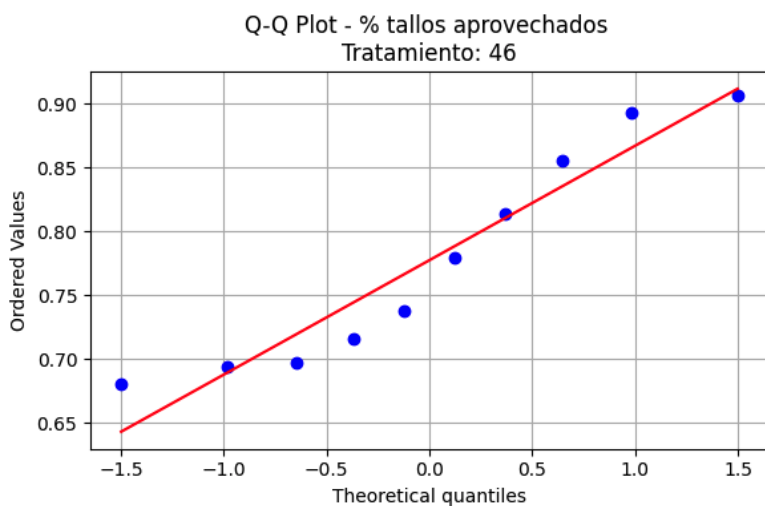
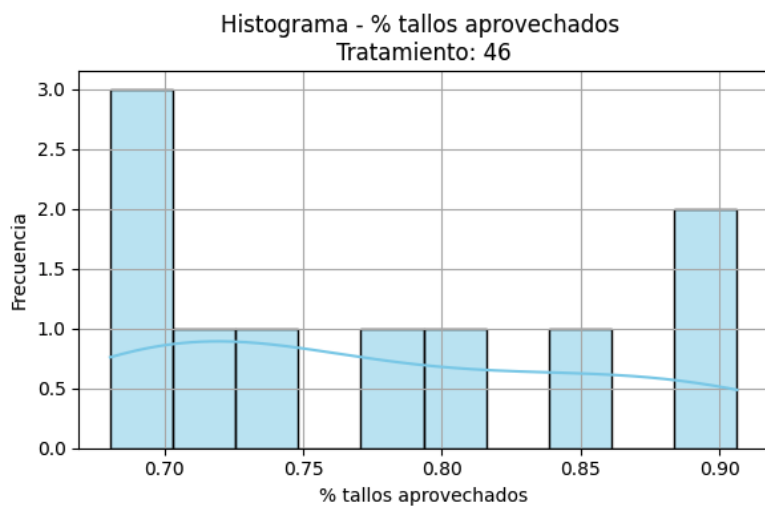
Variable: Peso por tallo
 Shapiro-Wilk: estadístico = 0.9126, p-valor = 0.2996
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 4.7052, p-valor = 0.0177
 Cumple con homocedasticidad: No

**Tratamiento T1***% Tallos Aprovechados*

El resultado de la aplicación de código para la variable % tallos aprovechados en el tratamiento T1. Cumple normalidad, no cumple homocedasticidad.

Figura 17*Prueba de Normalidad % Tallos de Aprovechamiento Tratamiento T1*

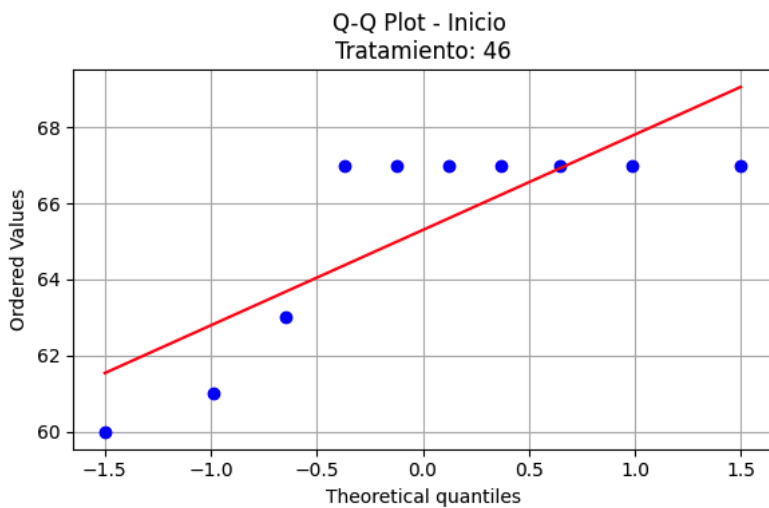
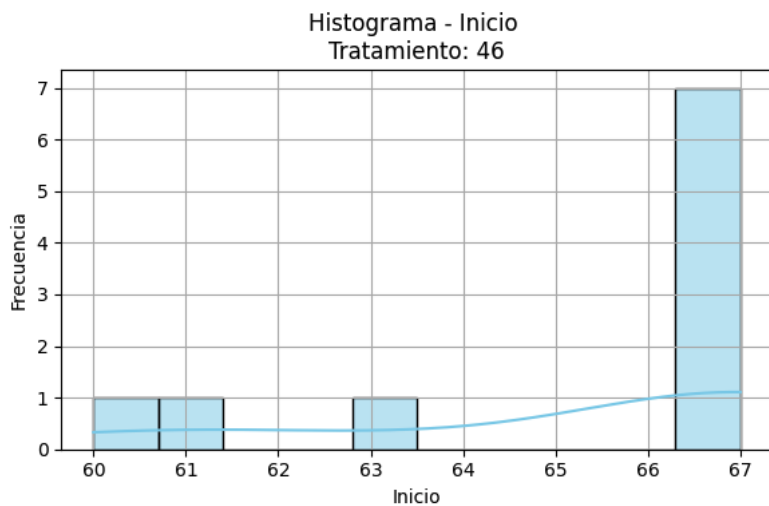
Variable: % tallos aprovechados
 Shapiro-Wilk: estadístico = 0.8983, p-valor = 0.2099
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 8.8961, p-valor = 0.0011
 Cumple con homocedasticidad: No

*Inicio de Cosecha.*

El resultado de la aplicación de código para la variable inicio de cosecha en el tratamiento T1. No cumple normalidad, Cumple homocedasticidad.

Figura 18*Prueba de Normalidad Inicio de Cosecha Tratamiento T1*

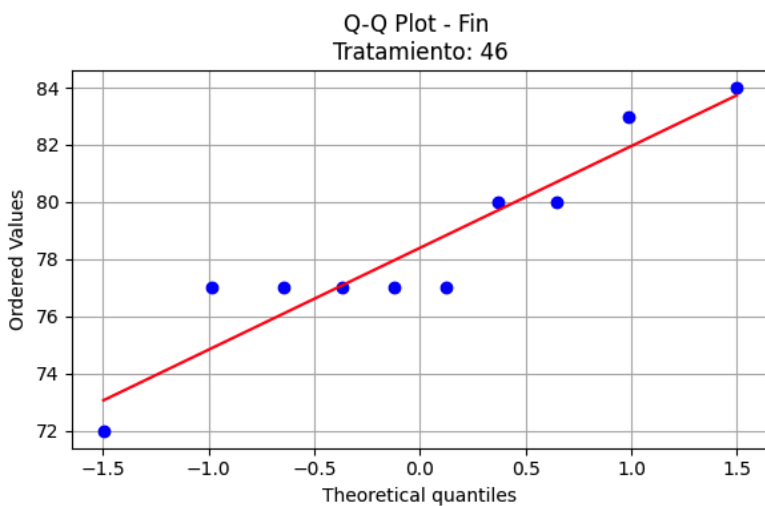
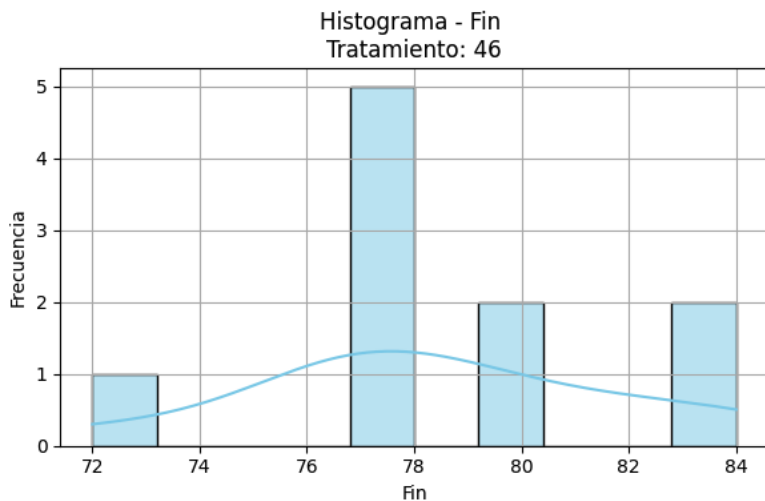
Variable: Inicio
 Shapiro-Wilk: estadístico = 0.6503, p-valor = 0.0002
 No es distribución normal (se rechaza H0)
 Prueba de Levene: estadístico = 0.7532, p-valor = 0.4805
 Cumple con homocedasticidad: Si

*Fin de Cosecha*

El resultado de la aplicación de código para la variable fin de cosecha en el tratamiento T1. Cumple normalidad, Cumple homocedasticidad.

Figura 19*Prueba de Normalidad Fin de Cosecha Tratamiento T1*

Variable: Fin
 Shapiro-Wilk: estadístico = 0.8986, p-valor = 0.2117
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 2.4334, p-valor = 0.1067
 Cumple con homocedasticidad: Sí

*Spam de Cosecha*

El resultado de la aplicación de código para la variable Spam de cosecha en el tratamiento T1. No cumple normalidad, Cumple homocedasticidad.

Figura 20

Prueba de Normalidad Spam de Cosecha Tratamiento T1

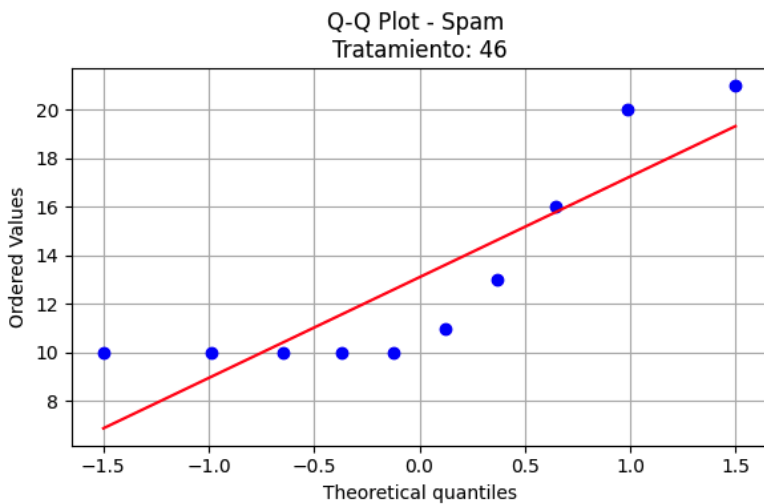
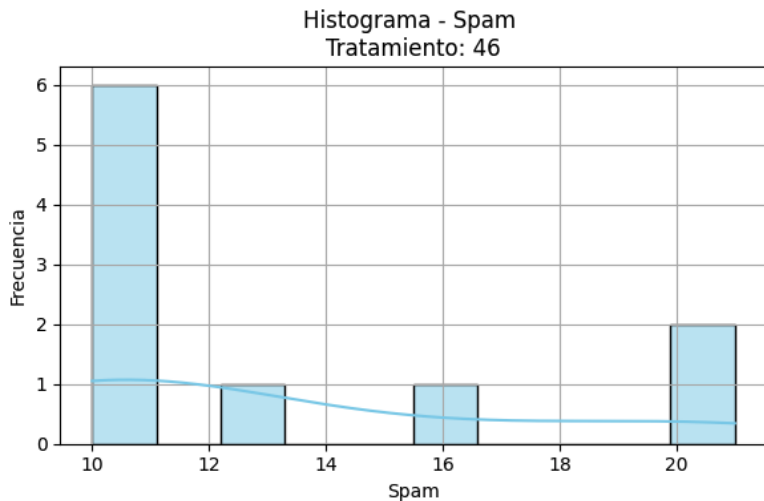
Variable: Spam

Shapiro-Wilk: estadístico = 0.7465, p-valor = 0.0032

No es distribución normal (se rechaza H0)

Prueba de Levene: estadístico = 0.7937, p-valor = 0.4624

Cumple con homocedasticidad: Sí

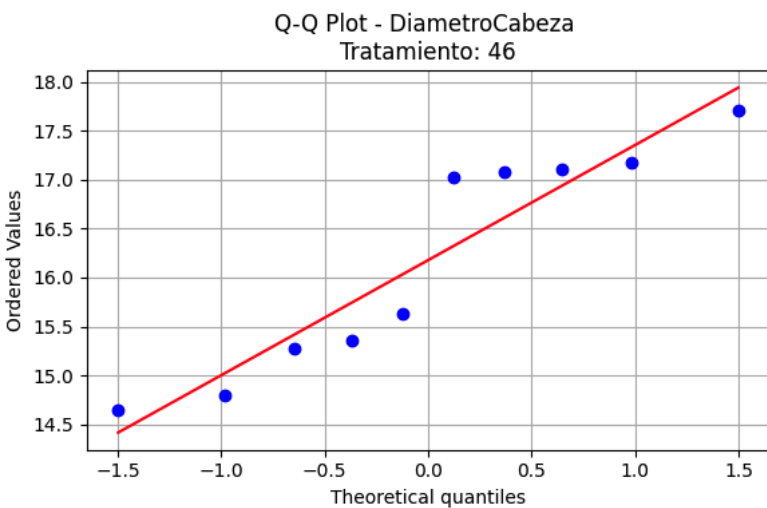
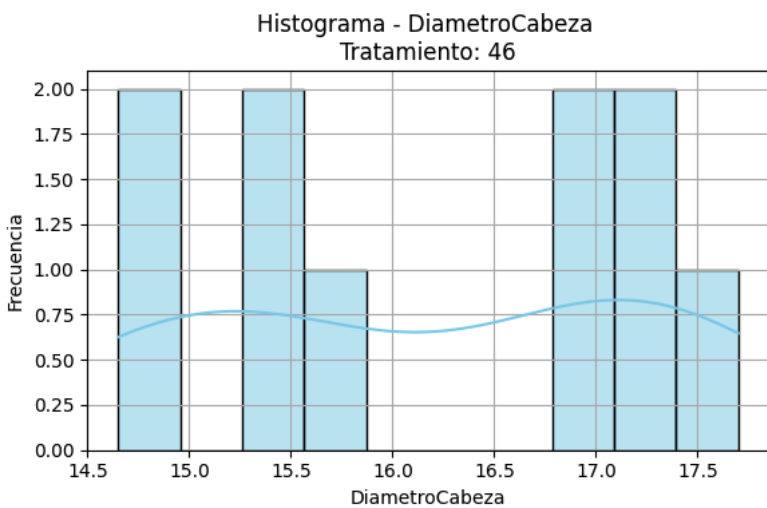


Diámetro de Cabeza

El resultado de la aplicación de código para la variable Diámetro de cabeza en el tratamiento T1. Cumple normalidad, No Cumple homocedasticidad.

Figura 21*Prueba de Normalidad Diámetro de Cabeza Tratamiento T1*

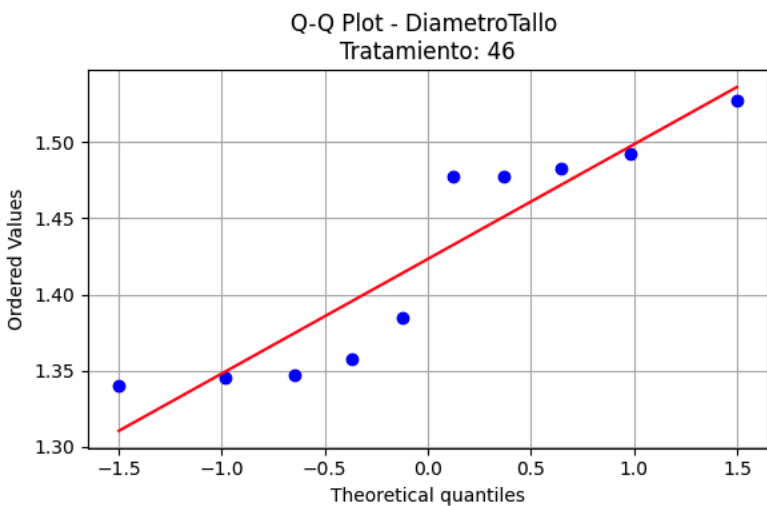
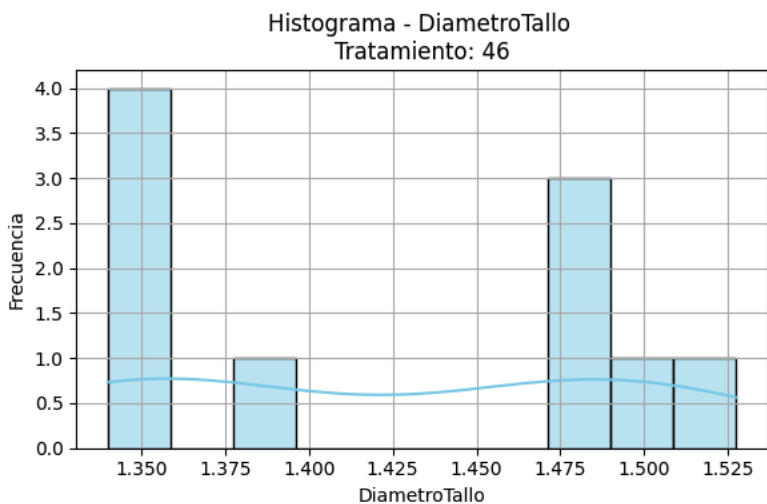
Variable: DiametroCabeza
 Shapiro-Wilk: estadístico = 0.8682, p-valor = 0.0952
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 18.1369, p-valor = 0.0000
 Cumple con homocedasticidad: No

*Diámetro de Tallo*

El resultado de la aplicación de código para la variable Diámetro de tallo en el tratamiento T1. No Cumple normalidad, Cumple homocedasticidad.

Figura 22*Prueba de Normalidad Diámetro de Tallo Tratamiento T1*

Variable: DiametroTallo
 Shapiro-Wilk: estadístico = 0.8341, p-valor = 0.0375
 No es distribución normal (se rechaza H0)
 Prueba de Levene: estadístico = 1.2655, p-valor = 0.2983
 Cumple con homocedasticidad: Sí

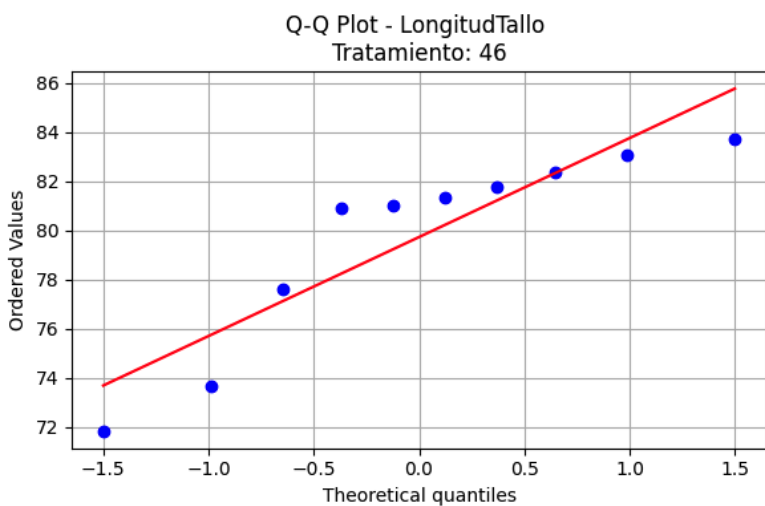
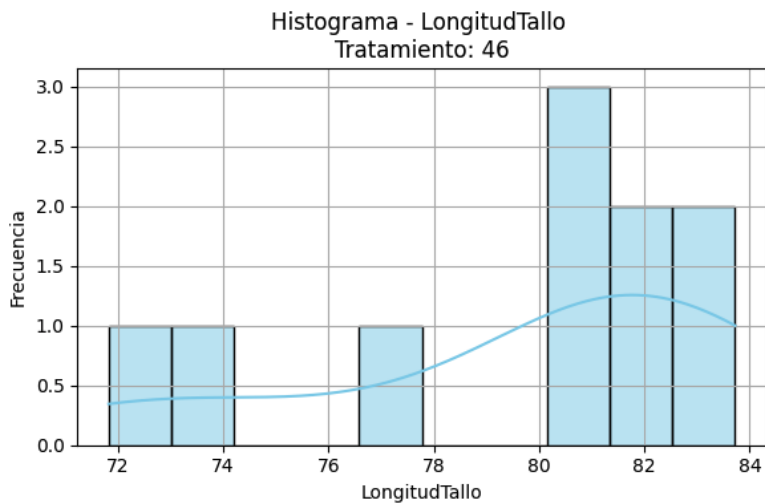
*Longitud de Tallo*

El resultado de la aplicación de código para la variable Longitud de tallo en el tratamiento T1. No Cumple normalidad, No Cumple homocedasticidad.

Figura 23

Prueba de Normalidad Longitud de Tallo Tratamiento T1

Variable: LongitudTallo
 Shapiro-Wilk: estadístico = 0.8297, p-valor = 0.0332
 No es distribución normal (se rechaza H0)
 Prueba de Levene: estadístico = 12.0861, p-valor = 0.0002
 Cumple con homocedasticidad: No



Peso por Tallo

El resultado de la aplicación de código para la variable Peso por tallo en el tratamiento

T1. Cumple normalidad, No Cumple homocedasticidad.

Figura 24

Prueba de Normalidad Peso por Tallo Tratamiento T1

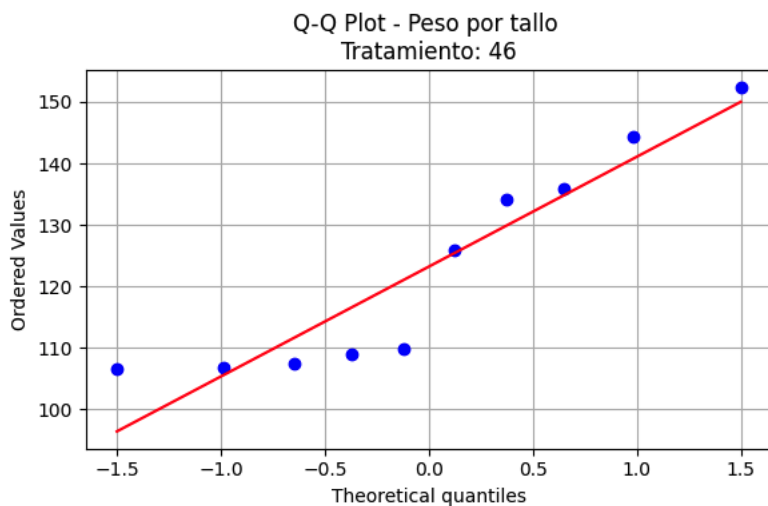
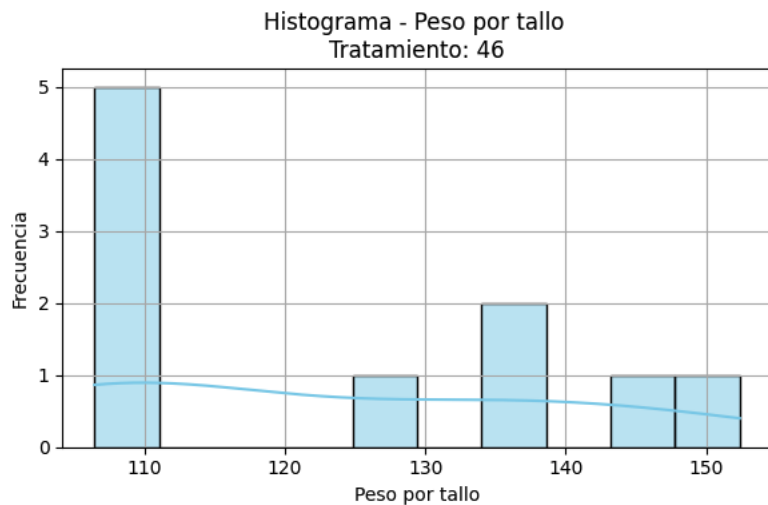
Variable: Peso por tallo

Shapiro-Wilk: estadístico = 0.8514, p-valor = 0.0603

Distribución normal (no se rechaza H0)

Prueba de Levene: estadístico = 4.7052, p-valor = 0.0177

Cumple con homocedasticidad: No



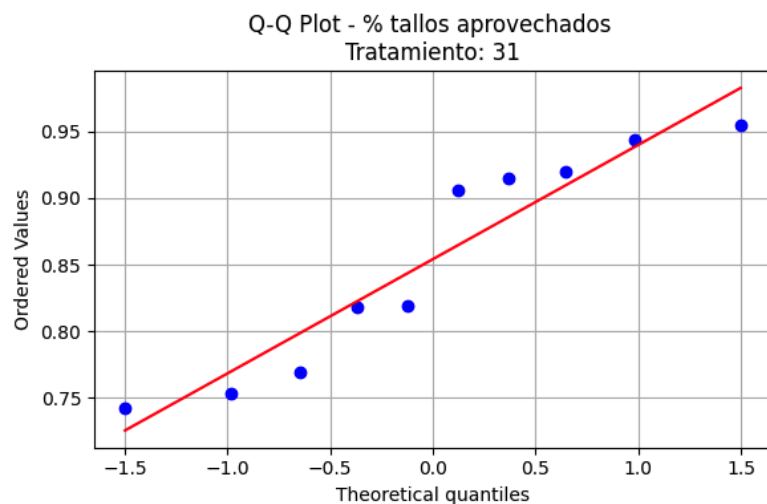
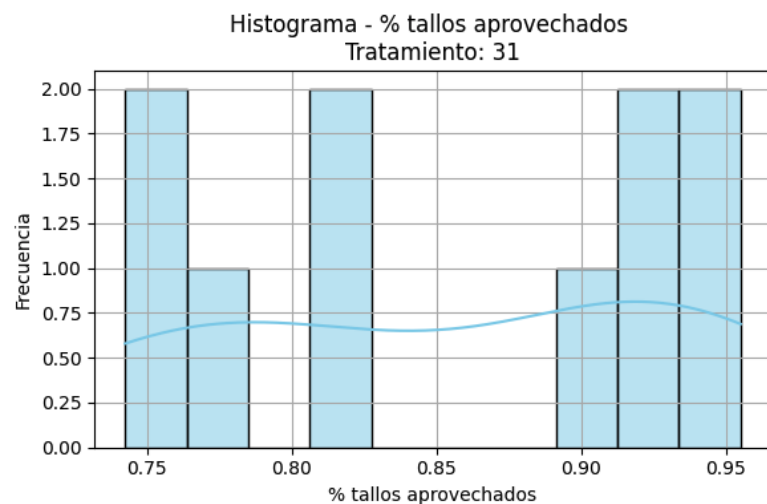
Tratamiento T2

% Tallos Aprovechados

El resultado de la aplicación de código para la variable % tallos aprovechados en el tratamiento T2. Cumple normalidad, No Cumple homocedasticidad.

Figura 25*Prueba de Normalidad % Tallos Aprovechados Tratamiento T2*

Variable: % tallos aprovechados
 Shapiro-Wilk: estadístico = 0.8798, p-valor = 0.1298
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 8.8961, p-valor = 0.0011
 Cumple con homocedasticidad: No

*Inicio de Cosecha*

El resultado de la aplicación de código para la variable Inicio de cosecha en el tratamiento T2. No Cumple normalidad, No Cumple homocedasticidad.

Figura 26

Prueba de Normalidad Inicio de Cosecha Tratamiento T2

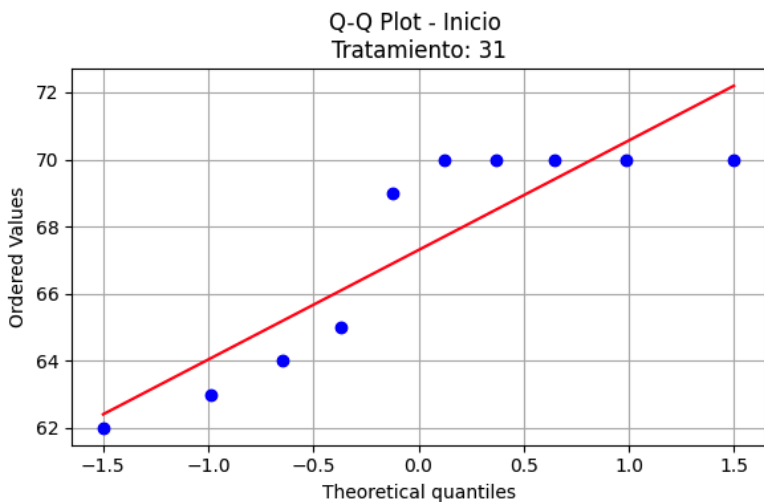
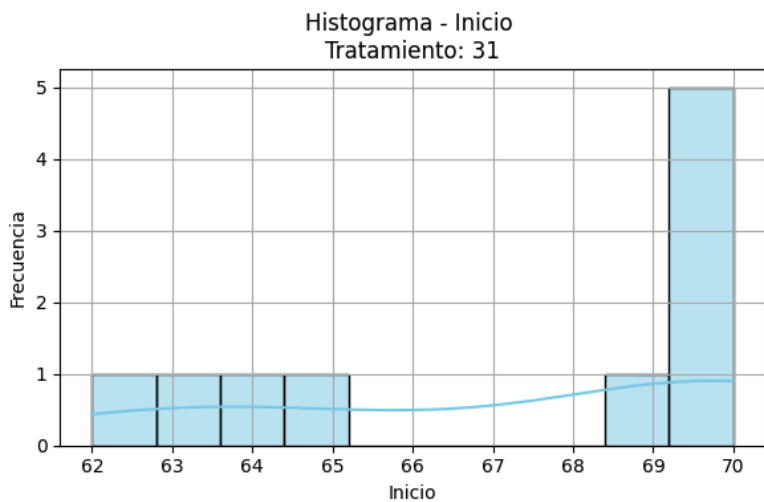
Variable: Inicio

Shapiro-Wilk: estadístico = 0.7641, p-valor = 0.0053

No es distribución normal (se rechaza H0)

Prueba de Levene: estadístico = 0.7532, p-valor = 0.4805

Cumple con homocedasticidad: Sí



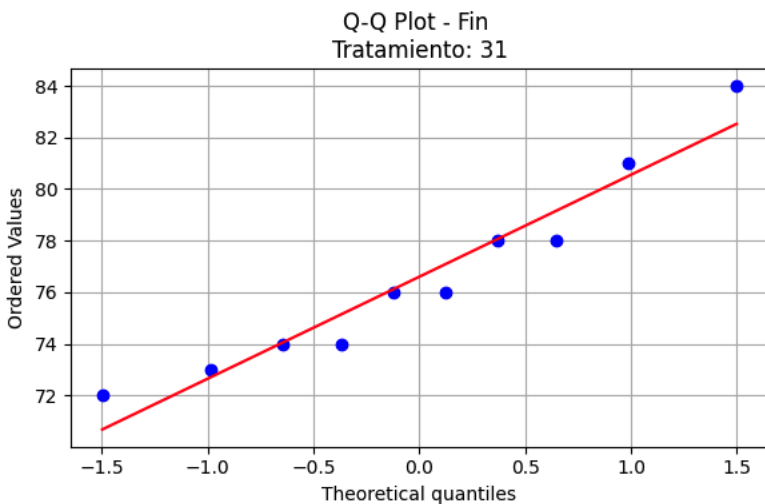
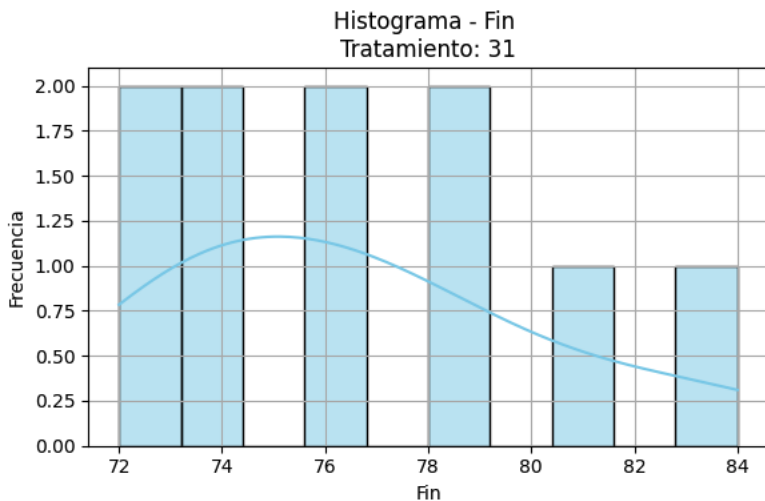
Fin de Cosecha

El resultado de la aplicación de código para la variable Fin de cosecha en el tratamiento

T2. No Cumple normalidad, Cumple homocedasticidad.

Figura 27*Prueba de Normalidad Fin de Cosecha Tratamiento T2*

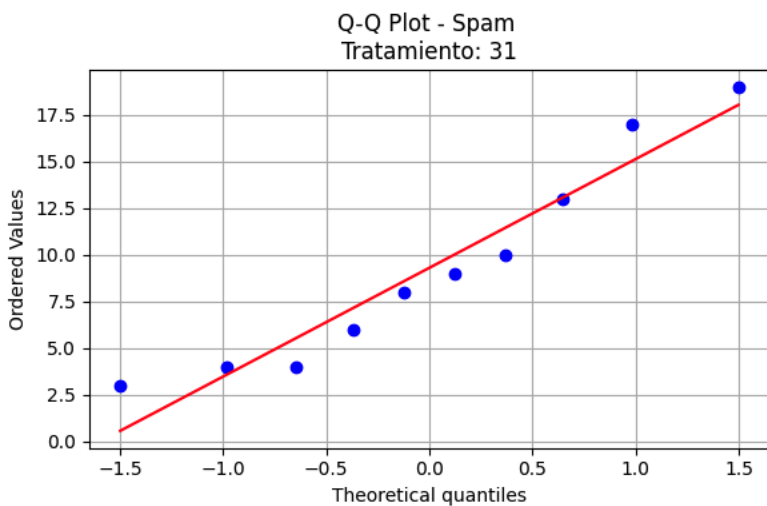
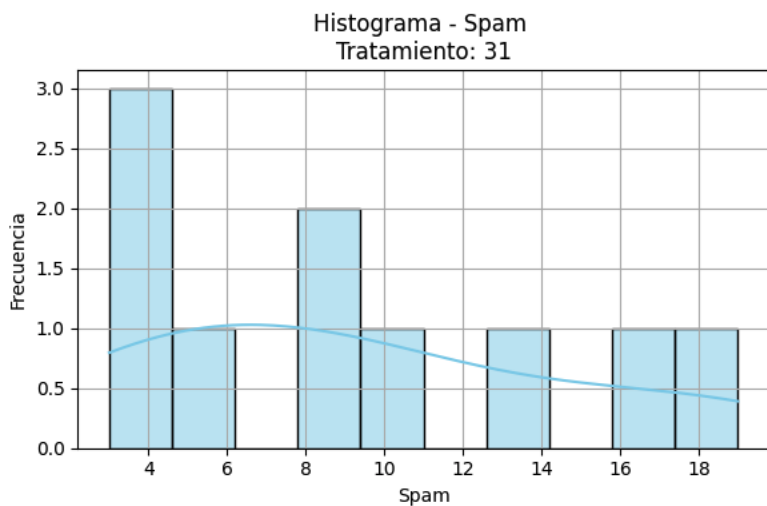
Variable: Fin
 Shapiro-Wilk: estadístico = 0.9339, p-valor = 0.4874
 Distribución normal (no se rechaza H₀)
 Prueba de Levene: estadístico = 2.4334, p-valor = 0.1067
 Cumple con homocedasticidad: Sí

*Spam de Cosecha*

El resultado de la aplicación de código para la variable Spam de cosecha en el tratamiento T2. Cumple normalidad, Cumple homocedasticidad.

Figura 28*Prueba de Normalidad Spam de Cosecha Tratamiento T2*

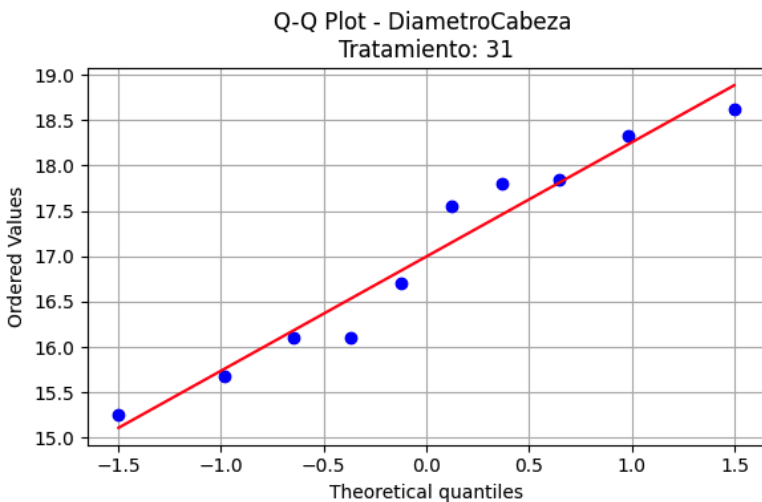
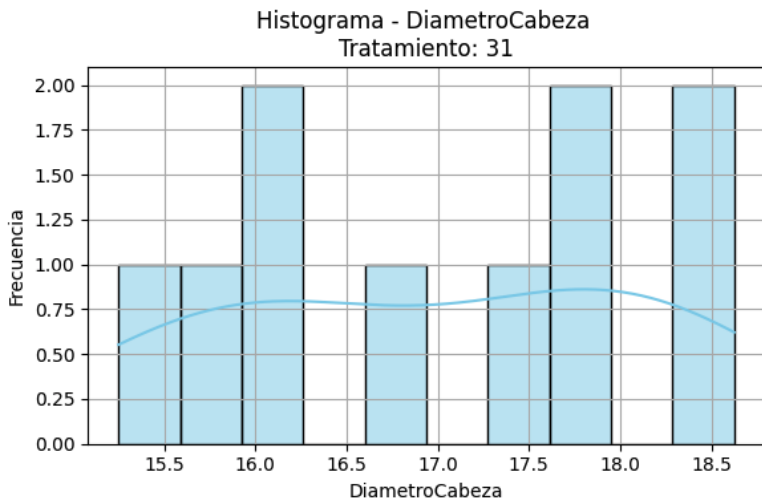
Variable: Spam
 Shapiro-Wilk: estadístico = 0.9175, p-valor = 0.3369
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 0.7937, p-valor = 0.4624
 Cumple con homocedasticidad: Sí

*Diámetro de Cabeza*

El resultado de la aplicación de código para la variable Diámetro de cabeza en el tratamiento T2. Cumple normalidad, No Cumple homocedasticidad.

Figura 29*Prueba de Normalidad Diámetro de Cabeza Tratamiento T2*

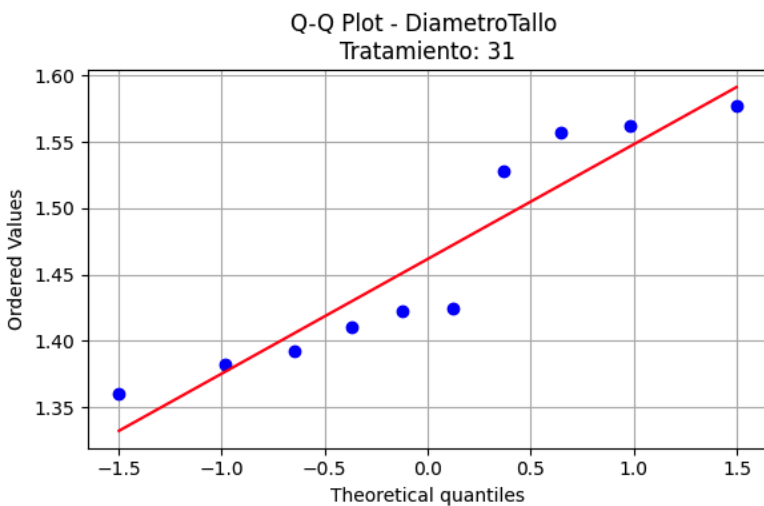
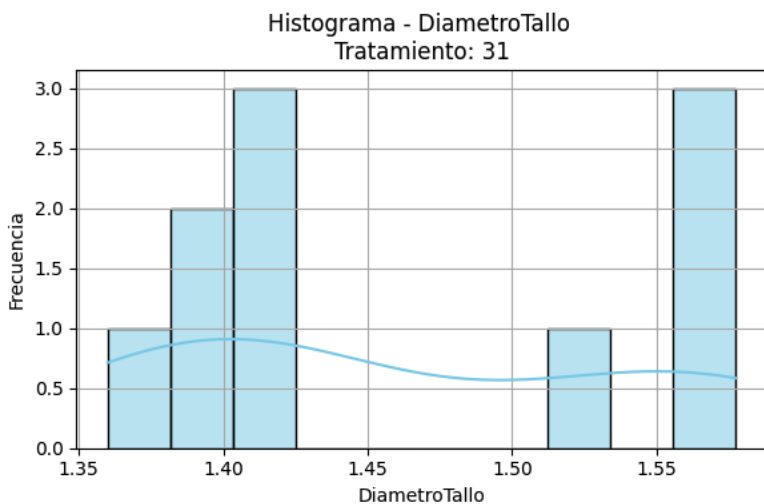
Variable: DiametroCabeza
 Shapiro-Wilk: estadístico = 0.9323, p-valor = 0.4712
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 18.1369, p-valor = 0.0000
 Cumple con homocedasticidad: No

*Diámetro de Tallo*

El resultado de la aplicación de código para la variable Diámetro de tallo en el tratamiento T2. Cumple normalidad, Cumple homocedasticidad.

Figura 30*Prueba de Normalidad Diámetro de Tallo Tratamiento T2*

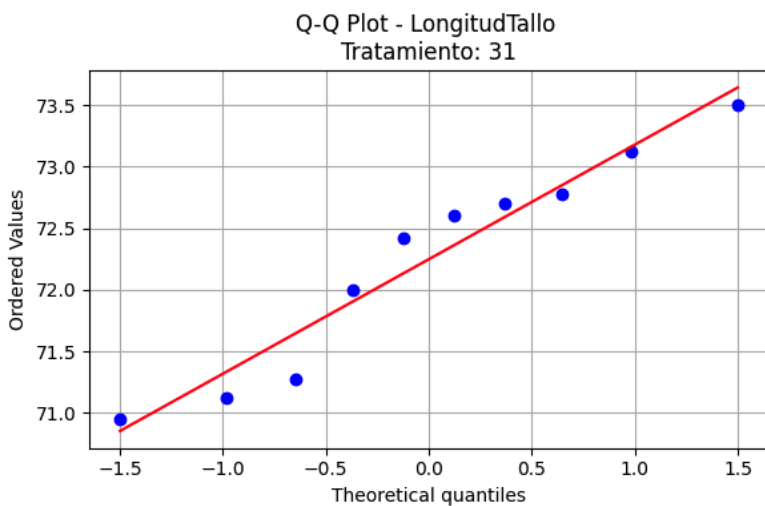
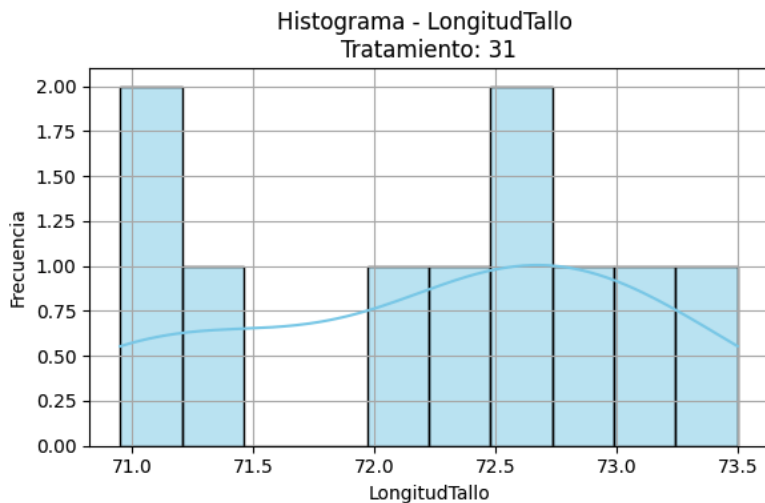
Variable: DiametroTallo
 Shapiro-Wilk: estadístico = 0.8543, p-valor = 0.0653
 Distribución normal (no se rechaza H₀)
 Prueba de Levene: estadístico = 1.2655, p-valor = 0.2983
 Cumple con homocedasticidad: Sí

*Longitud de Tallo*

El resultado de la aplicación de código para la variable Diámetro de tallo en el tratamiento T2. Cumple normalidad, No Cumple homocedasticidad.

Figura 31*Prueba de Normalidad Longitud de Tallo Tratamiento T2*

Variable: LongitudTallo
 Shapiro-Wilk: estadístico = 0.9279, p-valor = 0.4272
 Distribución normal (no se rechaza H0)
 Prueba de Levene: estadístico = 12.0861, p-valor = 0.0002
 Cumple con homocedasticidad: No

***Peso por Tallo***

El resultado de la aplicación de código para la variable Peso por tallo en el tratamiento

T2. Cumple normalidad, No Cumple homocedasticidad.

Figura 32*Prueba de Normalidad Peso por Tallo Tratamiento T2*

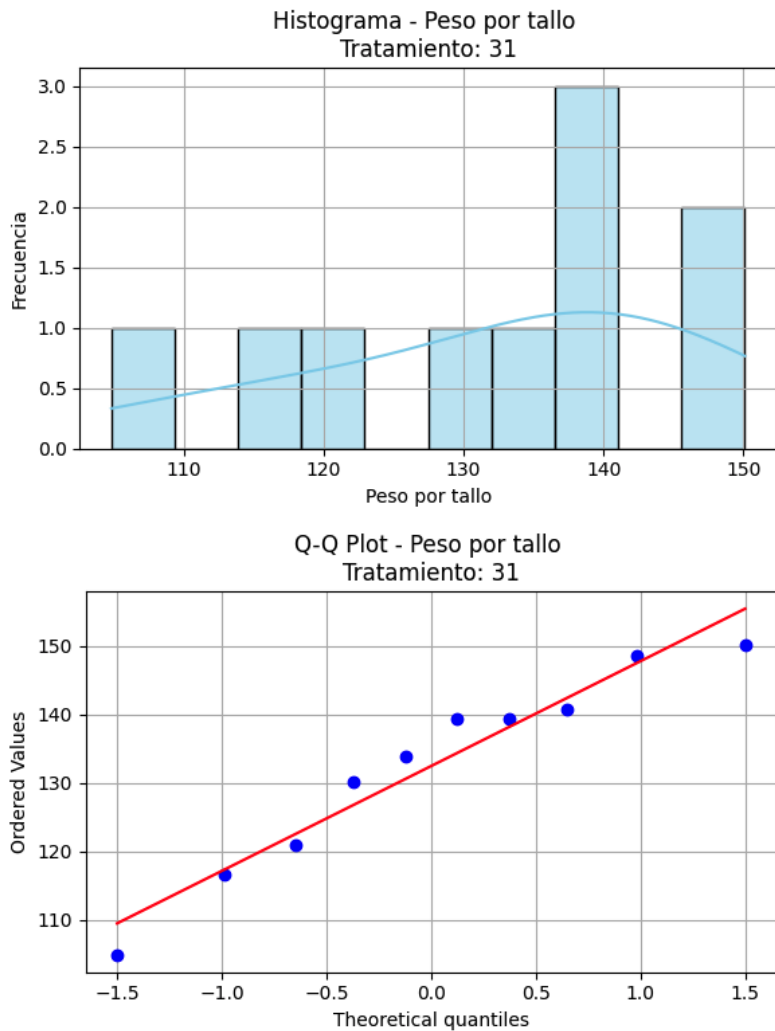
Variable: Peso por tallo

Shapiro-Wilk: estadístico = 0.9378, p-valor = 0.5289

Distribución normal (no se rechaza H₀)

Prueba de Levene: estadístico = 4.7052, p-valor = 0.0177

Cumple con homocedasticidad: No



En resumen, la prueba de Shapiro-Wilk que 6 de las variables de datos tomados en el ensayo no siguen una distribución normal.

Tabla 1*Resumen Shapiro-Wilk Variables por Tratamiento*

Variables/Tratamientos	T0 (57)	T1 (46)	T2 (31)
% tallos aprovechados	No (p=0.0115)	Sí (p=0.2099)	Sí (p=0.1298)
Inicio	Sí (p=0.1520)	No (p=0.0002)	No (p=0.0053)
Fin	Sí (p=0.2346)	Sí (p=0.2117)	Sí (p=0.4874)
Spam	Sí (p=0.1528)	No (p=0.0032)	Sí (p=0.3369)
DiametroCabeza	Sí (p=0.4482)	Sí (p=0.0952)	Sí (p=0.4712)
DiametroTallo	Sí (p=0.1224)	No (p=0.0375)	Sí (p=0.0653)
LongitudTallo	Sí (p=0.1103)	No (p=0.0332)	Sí (p=0.4272)
Peso por tallo	Sí (p=0.2996)	Sí (p=0.0603)	Sí (p=0.5289)

La prueba de Levene es una prueba estadística inferencial utilizada para evaluar la igualdad de las varianzas para una variable calculada para dos o más grupos, en este caso tenemos 4 variables: % tallos aprovechados, Diámetro de cabeza, Longitud de tallo y Peso por tallo cuya varianza no es igual entre los tratamientos.

Tabla 2*Resumen Levene Variables por Tratamiento*

Variables/Tratamientos	T0 (57)	T1 (46)	T2 (31)
% tallos aprovechados	No (p=0.0011)	No (p=0.0011)	No (p=0.0011)
Inicio	Sí (p=0.4805)	Sí (p=0.4805)	Sí (p=0.4805)
Fin	Sí (p=0.1067)	Sí (p=0.1067)	Sí (p=0.1067)
Spam	Sí (p=0.4624)	Sí (p=0.4624)	Sí (p=0.4624)

VARIABLES/TRATAMIENTOS	T0 (57)	T1 (46)	T2 (31)
DiametroCabeza	No (p=0.0000)	No (p=0.0000)	No (p=0.0000)
DiametroTallo	Sí (p=0.2983)	Sí (p=0.2983)	Sí (p=0.2983)
LongitudTallo	No (p=0.0002)	No (p=0.0002)	No (p=0.0002)
Peso por tallo	No (p=0.0177)	No (p=0.0177)	No (p=0.0177)

Matriz de Correlación de Kendall

Como primera medida para el análisis de los datos, realizamos un análisis de correlación entre las variables para identificar posibles relaciones y teniendo en cuenta que las variables no siguen un comportamiento normal, no usamos la matriz de correlación de Spearman, si no la matriz de correlación de Kendall.

A continuación, el código en Python para el cálculo de la matriz de correlación usando el método de Kendall, la matriz aplica a las variables numéricas, adicional se construye un mapa de calor.

Figura 33

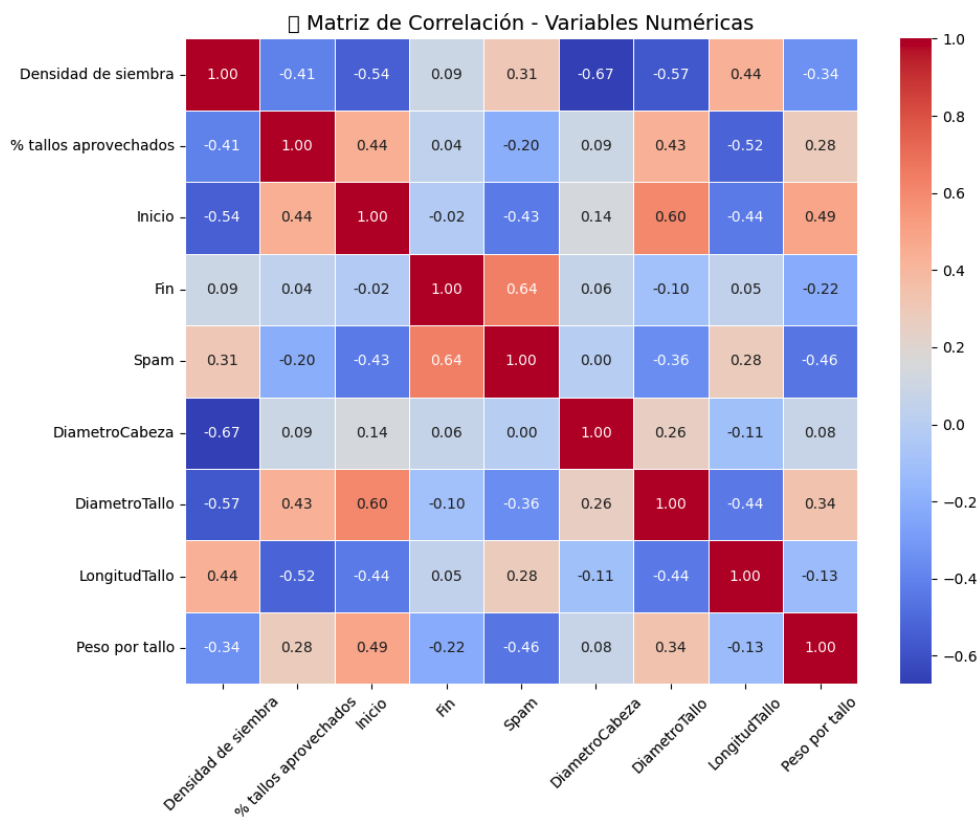
Código en Python Para Construir Matriz de Correlación Método Kendall

```
# Calcular matriz de correlación usando el método Kendall
matriz_correlacion_kendall = df1.select_dtypes(include=['float64', 'int64']).corr(method='kendall')

# Configurar el tamaño del gráfico
plt.figure(figsize=(10, 8))

# Dibujar el mapa de calor
sns.heatmap(matriz_correlacion_kendall, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)

# Título y ajustes
plt.title("🔗 Matriz de Correlación - Variables Numéricas", fontsize=14)
plt.xticks(rotation=45)
plt.yticks(rotation=0)
plt.tight_layout()
plt.show()
|
```

Figura 34*Matriz de Correlación de Kendall*

A continuación, la relación entre cada variable con base en los coeficientes de correlación de Pearson, destacando las asociaciones fuertes (positivas o negativas) y lo que podrían implicar. Correlaciones cercanas a 1 o -1 indican relaciones fuertes. Correlaciones cercanas a 0 indican relaciones débiles o nulas. Signo positivo indica relación directa, y signo negativo indica relación inversa

Densidad de Siembra

Negativamente correlacionada con: % tallos aprovechados (-0.41): A mayor densidad, menor aprovechamiento, posiblemente por competencia entre plantas. Inicio (-0.54): A mayor densidad, la cosecha inicia antes. Diámetro de cabeza (-0.67), Diámetro de tallo (-0.57):

Densidades altas reducen el tamaño de la planta, por menor espacio y nutrientes. Peso por tallo (-0.34): El peso individual disminuye con más plantas por área.

Positivamente correlacionada con: Longitud de tallo (0.44): Mayor densidad promueve tallos más largos, probablemente por competencia por luz. Spam (0.31): Se alarga ligeramente el periodo de cosecha.

En resumen, la Alta densidad disminuye tamaño y peso, pero puede adelantar y extender la cosecha.

% Tallos Aprovechados

Positivamente correlacionada con: Inicio (0.44): Inicios más tardíos permiten mayor aprovechamiento (mayor desarrollo). Diámetro de tallo (0.43): Tallos más gruesos se aprovechan más. Peso por tallo (0.28): Mayor peso, mejor aprovechamiento.

Negativamente correlacionada con: Longitud de tallo (-0.52): Tallos muy largos podrían no ser óptimos para cosecha. Spam (-0.20): Aprovechamientos altos tienden a ocurrir en periodos de cosecha más concentrados.

En resumen, el aprovechamiento está ligado a tallos más gruesos y pesados, con menos dispersión en la cosecha.

Inicio de Cosecha

Positivamente correlacionada con: Diámetro de tallo (0.60): Cosechas más tardías se asocian con tallos más gruesos. Peso por tallo (0.49): Más tiempo antes de cosechar permite mayor peso.

Negativamente correlacionada con: Spam (-0.43): Cosechas tempranas suelen durar más. Longitud de tallo (-0.44): Mayor longitud en cosechas más tempranas (posible efecto de estrés o competencia).

En resumen, inicios tardíos favorecen calidad (peso, grosor), mientras que los tempranos podrían extender la cosecha.

Fin de Cosecha

Positivamente correlacionada con: Spam (0.64): Un final de cosecha más tardío está asociado con una duración más larga.

Correlaciones muy débiles o nulas: Inicio (-0.02), % aprovechamiento (0.04), Densidad (0.09)

En resumen, el final de la cosecha se relaciona principalmente con la duración total (Spam); las demás relaciones son irrelevantes.

Spam (duración de cosecha)

Positivamente correlacionada con: Fin (0.64): Cosechas que terminan más tarde suelen tener un spam de corte mucho más largo.

Negativamente correlacionada con: Inicio (-0.43): Cuanto más tarde inicia, más corta es la duración. Peso por tallo (-0.46): Cosechas más largas tienden a tener tallos más livianos. Diámetro de tallo (-0.36): Ligera tendencia a tallos más delgados.

En resumen, cosechas largas están relacionadas con calidad ligeramente inferior (peso, grosor).

Diámetro de Cabeza

Negativamente correlacionada con: Densidad de siembra (-0.67): A mayor densidad, cabezas más pequeñas, por competencia. Longitud del tallo (-0.11): Tallos más largos tienden a tener cabezas levemente más pequeñas.

Positivamente correlacionada con: Inicio (0.14), % aprovechamiento (0.09), Diámetro de tallo (0.26), Peso por tallo (0.08): Correlaciones débiles.

En resumen, el diámetro de la cabeza se reduce con alta densidad. Las demás relaciones son débiles o poco informativas

Diámetro del Tallo

Negativamente correlacionada con: Densidad (-0.57): Más densidad reduce el grosor del tallo. Spam (-0.36), Fin (-0.09): Ligeramente disminución del diámetro con cosechas más prolongadas o tardías. Longitud del tallo (-0.44): Tallos más largos tienden a ser más delgados.

Positivamente correlacionada con: Inicio (0.60): Mayor tiempo antes de cosecha permite mayor engrosamiento. % aprovechamiento (0.43), Peso por tallo (0.33): Mejores rendimientos se asocian con mayor diámetro.

En resumen: Tallos más gruesos aparecen en densidades bajas y cosechas más tardías; además, están ligados a mayor aprovechamiento y peso.

Longitud del Tallo

Positivamente correlacionada con: Densidad (0.44): A mayor densidad, más longitud (búsqueda de luz). Spam (0.28): Cosechas más largas tienen tallos ligeramente más largos.

Negativamente correlacionada con: % aprovechamiento (-0.52), Inicio (-0.44), Diámetro de tallo (-0.44): Tallos largos no necesariamente son aprovechables ni gruesos. Peso por tallo (-0.13): Relación débilmente negativa.

En resumen los tallos largos surgen en densidades altas y cosechas prolongadas, pero no implican mejor calidad.

Peso por Tallo

Negativamente correlacionada con: Densidad (-0.34): A mayor densidad, menos peso individual. Spam (-0.46): Cosechas más largas tienden a tallos más livianos. Longitud del tallo (-0.13): Tallos más largos no significan más peso.

Positivamente correlacionada con: Inicio (0.49): Más tiempo antes de la cosecha incrementa el peso. Diámetro de tallo (0.33), % aprovechamiento (0.28): Tallos más pesados suelen ser más aprovechables y gruesos.

En resumen, el peso de los tallos aumenta con tiempo y espacio, y se asocia con calidad (mayor aprovechamiento y diámetro).

La densidad de siembra influye de forma negativa en la calidad del producto individual, pero modifica significativamente el comportamiento del cultivo a lo largo del tiempo. A medida que se incrementa la densidad, se observa una disminución en el porcentaje de tallos aprovechados, así como en el diámetro de la cabeza, el diámetro del tallo y el peso por tallo, lo que sugiere una mayor competencia entre plantas por luz, nutrientes y espacio, reduciendo así la calidad y el aprovechamiento comercial por unidad. Además, la densidad más alta adelanta el inicio de la cosecha y tiende a alargar el intervalo entre el inicio y el fin de esta, lo cual puede resultar útil para estrategias de cosecha escalonada. Sin embargo, no se evidencia una relación clara entre la densidad y la fecha de finalización de la cosecha, lo que indica que el efecto se concentra principalmente en el inicio del ciclo. También se nota una elongación del tallo en condiciones de mayor densidad, posiblemente como una respuesta adaptativa a la búsqueda de luz, fenómeno conocido como etiolación. En conjunto, estos hallazgos resaltan la necesidad de considerar cuidadosamente la densidad de siembra como una variable estratégica que impacta tanto la calidad del producto como el ciclo del cultivo.

Después de analizar la matriz de correlación y evidenciar relaciones significativas entre las variables, damos el siguiente paso en el análisis estadístico: la aplicación de un análisis de varianza (Kruskal-Wallis) para determinar si estas asociaciones son estadísticamente significativas.

Kruskal-Wallis

Dado que varias de las variables evaluadas en el experimento no cumplen con el supuesto de normalidad según la prueba de Shapiro-Wilk, y considerando además que algunas presentan heterogeneidad de varianzas entre tratamientos (según la prueba de Levene), el uso de métodos paramétricos como el ANOVA no resulta adecuado. Por lo tanto, se optó por aplicar la prueba de Kruskal-Wallis, una alternativa no paramétrica que permite comparar la mediana de los grupos sin asumir normalidad ni homogeneidad de varianzas.

Esta prueba resulta especialmente útil en contextos experimentales como el presente, donde se desea determinar si existen diferencias estadísticamente significativas en las variables de respuesta (como diámetro de cabeza, longitud del tallo, o porcentaje de tallos aprovechados) entre los distintos niveles de la variable independiente (densidad de siembra), sin que los datos cumplan estrictamente los supuestos clásicos del análisis de varianza.

Figura 35

Código en Python Para Aplicar Prueba de Kruskal-Wallis

```
from scipy.stats import kruskal

# Filtrar columnas numéricas
columnas_numericas = df1.select_dtypes(include=['float64', 'int64']).columns

# Recorrer variables
for columna in columnas_numericas:
    try:
        grupo_T0 = df1[df1['Tratamiento'] == 'T0'][columna]
        grupo_T1 = df1[df1['Tratamiento'] == 'T1'][columna]
        grupo_T2 = df1[df1['Tratamiento'] == 'T2'][columna]

        stat, p = kruskal(grupo_T0, grupo_T1, grupo_T2)

        print(f"\n Variable: {columna}")
        print(f"   Estadístico de Kruskal-Wallis = {stat:.4f}, p-valor = {p:.4f}")
        if p < 0.05:
            print("       Hay diferencias significativas entre tratamientos.")
        else:
            print("       No hay diferencias significativas entre tratamientos.")
    except Exception as e:
        print(f"\n Error al procesar la variable '{columna}': {e}")
```

En resumen, las variables que siguen un comportamiento normal según la prueba de shapiro Wilk son: Fin de corte, Spam, diámetro del tallo.

Figura 36*Resultado Prueba de Kruskal-Wallis*

Variable: Densidad de siembra
 Estadístico de Kruskal-Wallis = 29.0000, p-valor = 0.0000
 Hay diferencias significativas entre tratamientos.

Variable: % tallos aprovechados
 Estadístico de Kruskal-Wallis = 7.0838, p-valor = 0.0290
 Hay diferencias significativas entre tratamientos.

Variable: Inicio
 Estadístico de Kruskal-Wallis = 11.4194, p-valor = 0.0033
 Hay diferencias significativas entre tratamientos.

Variable: Fin
 Estadístico de Kruskal-Wallis = 0.8488, p-valor = 0.6542
 No hay diferencias significativas entre tratamientos.

Variable: Spam
 Estadístico de Kruskal-Wallis = 4.5613, p-valor = 0.1022
 No hay diferencias significativas entre tratamientos.

Variable: DiametroCabeza
 Estadístico de Kruskal-Wallis = 20.5157, p-valor = 0.0000
 Hay diferencias significativas entre tratamientos.

Variable: DiametroTallo
 Estadístico de Kruskal-Wallis = 15.2903, p-valor = 0.0005
 Hay diferencias significativas entre tratamientos.

Variable: LongitudTallo
 Estadístico de Kruskal-Wallis = 10.0335, p-valor = 0.0066
 Hay diferencias significativas entre tratamientos.

Variable: Peso por tallo
 Estadístico de Kruskal-Wallis = 6.4826, p-valor = 0.0391
 Hay diferencias significativas entre tratamientos.

Con base en los resultados obtenidos mediante la prueba de Kruskal-Wallis, se evidencia que existen diferencias estadísticamente significativas entre tratamientos para la mayoría de las variables evaluadas. Específicamente, variables como % de tallos aprovechados ($p = 0.0290$), inicio de cosecha ($p = 0.0033$), diámetro de cabeza ($p = 0.0000$), diámetro del tallo ($p = 0.0005$), longitud del tallo ($p = 0.0066$) y peso por tallo ($p = 0.0391$) presentan diferencias significativas, lo que indica que la densidad de siembra influye de manera diferenciada sobre estas respuestas

morfo-productivas. Por el contrario, variables como el fin de cosecha ($p = 0.6542$) y la duración del periodo de cosecha –Spam– ($p = 0.1022$) no mostraron diferencias estadísticamente significativas entre tratamientos, lo que sugiere que estos aspectos del ciclo de producción no se ven claramente modificados por los niveles de densidad evaluados. En conjunto, estos resultados refuerzan la importancia del manejo adecuado de la densidad de siembra como factor determinante en la calidad y comportamiento fenológico del cultivo.

Post Hoc de Dunn

Luego de aplicar la prueba de Kruskal-Wallis y detectar diferencias significativas entre tratamientos en varias de las variables analizadas, se procede a realizar una prueba post hoc de Dunn. Esta prueba permite identificar específicamente entre qué pares de tratamientos se presentan dichas diferencias, proporcionando así un análisis más detallado de los efectos del factor evaluado.

Figura 37

Código en Python Para Aplicar Prueba Post Hoc de Dunn

```

# Columnas numéricas excluyendo 'Fin' y 'Spam'
columnas_numericas = df1.select_dtypes(include=['float64', 'int64']).columns.drop(['Fin', 'Spam'])

# Comparaciones que se evaluarán
pares = [('T0', 'T1'), ('T0', 'T2'), ('T1', 'T2')]
for columna in columnas_numericas:
    try:
        print(f"\n Post hoc de Dunn para la variable: {columna}")

        # Aplicar prueba de Dunn con corrección de holm
        dunn_result = sp.posthoc_dunn(
            df1,
            val_col=columna,
            group_col='Tratamiento',
            p_adjust='holm'
        ).round(3)

        # Mostrar tabla completa
        print("\nTabla de p-valores:")
        print(dunn_result)

        # Interpretar cada comparación
        print("\n Interpretación por comparación:")
        for grupo1, grupo2 in pares:
            p_valor = dunn_result.loc[grupo1, grupo2]
            if p_valor < 0.05:
                print(f" Hay diferencia significativa entre {grupo1} y {grupo2} (p = {p_valor:.4f})")
            else:
                print(f" No hay diferencia significativa entre {grupo1} y {grupo2} (p = {p_valor:.4f})")

    except Exception as e:
        print(f"\n Error al procesar la variable '{columna}': {e}")

```

Figura 38

Resultado Prueba Post Hoc de Dunn Variable % Tallos Aprovechados

Post hoc de Dunn para la variable: % tallos aprovechados

Tabla de p-valores:

	T0	T1	T2
T0	1.000	0.409	0.028
T1	0.409	1.000	0.151
T2	0.028	0.151	1.000

Interpretación por comparación:

No hay diferencia significativa entre T0 y T1 (p = 0.4090)

Hay diferencia significativa entre T0 y T2 (p = 0.0280)

No hay diferencia significativa entre T1 y T2 (p = 0.1510)

Para la variable % de tallos aprovechados, encontramos diferencia estadística significativa entre el tratamiento T0 y T2

Figura 39

Resultado Prueba Post Hoc de Dunn Variable % Tallos Aprovechados

Post hoc de Dunn para la variable: Inicio

Tabla de p-valores:

	T0	T1	T2
T0	1.000	0.072	0.002
T1	0.072	1.000	0.212
T2	0.002	0.212	1.000

Interpretación por comparación:

No hay diferencia significativa entre T0 y T1 ($p = 0.0720$)

Hay diferencia significativa entre T0 y T2 ($p = 0.0020$)

No hay diferencia significativa entre T1 y T2 ($p = 0.2120$)

Figura 40

Resultado Prueba Post Hoc de Dunn Variable Diámetro de Cabezas

Post hoc de Dunn para la variable: DiametroCabeza

Tabla de p-valores:

	T0	T1	T2
T0	1.000	0.002	0.000
T1	0.002	1.000	0.286
T2	0.000	0.286	1.000

Interpretación por comparación:

Hay diferencia significativa entre T0 y T1 ($p = 0.0020$)

Hay diferencia significativa entre T0 y T2 ($p = 0.0000$)

No hay diferencia significativa entre T1 y T2 ($p = 0.2860$)

Figura 41

Resultado Prueba Post Hoc de Dunn Variable Diámetro de Tallo

Post hoc de Dunn para la variable: DiametroTallo

Tabla de p-valores:

	T0	T1	T2
T0	1.000	0.015	0.000
T1	0.015	1.000	0.253
T2	0.000	0.253	1.000

Interpretación por comparación:

Hay diferencia significativa entre T0 y T1 ($p = 0.0150$)

Hay diferencia significativa entre T0 y T2 ($p = 0.0000$)

No hay diferencia significativa entre T1 y T2 ($p = 0.2530$)

Figura 42

Resultado Prueba Post Hoc de Dunn Variable Longitud de Tallos

Post hoc de Dunn para la variable: LongitudTallo

Tabla de p-valores:

	T0	T1	T2
T0	1.000	1.000	0.018
T1	1.000	1.000	0.018
T2	0.018	0.018	1.000

Interpretación por comparación:

No hay diferencia significativa entre T0 y T1 ($p = 1.0000$)

Hay diferencia significativa entre T0 y T2 ($p = 0.0180$)

Hay diferencia significativa entre T1 y T2 ($p = 0.0180$)

Figura 43

Resultado Prueba Post Hoc de Dunn Variable Peso por Tallo

Post hoc de Dunn para la variable: Peso por tallo

Tabla de p-valores:

	T0	T1	T2
T0	1.000	0.31	0.033
T1	0.310	1.00	0.310
T2	0.033	0.31	1.000

Interpretación por comparación:

No hay diferencia significativa entre T0 y T1 (p = 0.3100)

Hay diferencia significativa entre T0 y T2 (p = 0.0330)

No hay diferencia significativa entre T1 y T2 (p = 0.3100)

Figura 44

Código en Python Para Calcular Medias de Cada Variable por Tratamiento

```
# Calcula las medianas
df_medianas = df1.groupby('Tratamiento')[columnas_numericas].median()

# Muestra el resultado
print(df_medianas)
```

Figura 45*Resultado Media por Variable por Tratamiento*

Tratamiento	Densidad de siembra	% tallos aprovechados	Inicio \	
T0	57.0	0.718528	62.0	
T1	46.0	0.758453	67.0	
T2	31.0	0.862699	69.5	

Tratamiento	DiametroCabeza	DiametroTallo	LongitudTallo	Peso por tallo
T0	13.400	1.23625	82.7500	112.5050
T1	16.325	1.43125	81.1500	117.8510
T2	17.125	1.42375	72.5125	136.5504

Con base en los resultados de la prueba post hoc de Dunn y las medianas por tratamiento, se puede detallar e interpretar de manera más profunda las diferencias significativas detectadas en las variables analizadas.

Para % de tallos aprovechados, se observó una diferencia significativa entre los tratamientos T0 y T2 ($p = 0.028$). La mediana fue menor en T0 (0.7185) y mayor en T2 (0.8627), lo cual indica que a menor densidad (T2), se logra un mayor aprovechamiento de tallos, probablemente por menor competencia intraespecífica.

En la variable Inicio de cosecha, se encontró una diferencia significativa entre T0 y T2 ($p = 0.002$), con medianas de 62.0 días para T0 y 69.5 días para T2. Esto sugiere que a menor densidad, el inicio de cosecha se retrasa, lo que es coherente con la hipótesis de que en densidades bajas, las plantas tardan más en alcanzar el estado de cosecha debido a menor competencia.

En cuanto al diámetro de cabeza, hubo diferencias significativas entre T0 y T1 ($p = 0.002$) y entre T0 y T2 ($p = 0.000$). Las medianas muestran un aumento progresivo desde T0

(13.4 cm) a T2 (17.1 cm), lo que confirma que a menor densidad, el diámetro de la cabeza incrementa significativamente, mejorando la calidad comercial del producto.

Respecto al diámetro del tallo, también se hallaron diferencias significativas entre T0 y T1 ($p = 0.015$) y entre T0 y T2 ($p = 0.000$). T0 presentó la mediana más baja (1.236 cm), mientras que T1 y T2 tuvieron valores similares (1.431 cm y 1.423 cm, respectivamente), lo que indica que densidades menores favorecen un tallo más grueso, aunque la diferencia entre T1 y T2 no fue estadísticamente significativa.

Para la longitud del tallo, se encontraron diferencias entre T0 y T2 ($p = 0.018$) y entre T1 y T2 ($p = 0.018$). Curiosamente, T2 mostró la menor longitud del tallo (72.5 cm), mientras que T0 y T1 tuvieron longitudes mayores (82.8 cm y 81.1 cm, respectivamente). Esto sugiere que a mayor densidad, los tallos tienden a alargarse, posiblemente como respuesta a la competencia por luz (etiolación).

Finalmente, en la variable peso por tallo, se detectó una diferencia significativa entre T0 y T2 ($p = 0.033$). El tratamiento T2 presentó la mediana más alta (136.6 g), en comparación con T0 (112.5 g), lo que indica que en densidades bajas el peso individual por planta es mayor, favoreciendo la producción por unidad.

En resumen, los resultados muestran que densidades más bajas (especialmente T2) favorecen el aprovechamiento, el tamaño (diámetro de cabeza y tallo) y el peso de los tallos, aunque retrasan el inicio de la cosecha. Estos hallazgos son claves para la toma de decisiones agronómicas según el objetivo del productor: maximizar calidad individual o sincronizar el ciclo productivo.

Conclusiones

El estudio evidenció que la densidad de siembra influye de manera significativa sobre diversas variables morfológicas y productivas en el cultivo de Brassicas. Dado que varias variables no cumplieron con los supuestos de normalidad (Shapiro-Wilk) ni homogeneidad de varianzas (Levene), se optó por un enfoque estadístico no paramétrico (Kruskal-Wallis y Dunn), lo que permitió una evaluación más rigurosa y adaptada a las características de los datos. La programación en Python facilitó un análisis ágil, reproducible y visualmente claro, lo cual fortaleció la toma de decisiones basada en evidencia técnica y estadística.

Respecto al diámetro de la flor, se identificó una correlación negativa fuerte con la densidad de siembra (-0.78), lo que indica que densidades más bajas favorecen un mayor desarrollo floral. Este resultado fue consistente con los hallazgos de Kruskal-Wallis y la prueba post hoc de Dunn, que detectaron diferencias significativas entre tratamientos, en especial entre T0 y T2. A pesar de que la normalidad se cumplió en todos los grupos, la prueba de Levene indicó heterocedasticidad ($p=0.0000$), justificando así el uso de pruebas no paramétricas. Este comportamiento sugiere que una menor competencia entre plantas permite un mayor desarrollo del órgano floral.

En cuanto al peso por tallo, se observó una correlación negativa con la densidad de siembra (-0.51) y positiva con el diámetro del tallo (0.57), lo que revela que en condiciones de menor competencia la biomasa individual tiende a incrementarse. Aunque los datos cumplieron con la normalidad en los tres tratamientos, la prueba de Levene mostró desigualdad de varianzas ($p=0.0177$), motivo por el cual se utilizó nuevamente Kruskal-Wallis. Los resultados mostraron diferencias significativas, destacando a T2 como el tratamiento con mayor peso por tallo, lo cual

puede ser beneficioso en mercados que priorizan calidad y tamaño individual sobre el volumen total.

El porcentaje de tallos aprovechados también mostró una tendencia favorable con la reducción en la densidad, evidenciada por su correlación negativa con esta variable (-0.52). Desde el punto de vista estadístico, la variable presentó heterocedasticidad y falta de normalidad en el tratamiento T0 (Levene: $p=0.0011$; Shapiro-Wilk: $p=0.0115$). El análisis no paramétrico reveló diferencias significativas, y el test de Dunn destacó a T2 como el tratamiento con mayor porcentaje aprovechado. Este hallazgo indica que, si bien se reduce la cantidad de plantas sembradas, se incrementa la proporción de aquellas que cumplen con los estándares de cosecha, optimizando así el rendimiento útil del cultivo.

En relación con la variable "Inicio" de la cosecha, se observó una correlación negativa con la densidad de siembra (-0.58), lo que implica que densidades menores retrasan ligeramente el comienzo del periodo de recolección. Aunque esta variable presentó homocedasticidad (Levene: $p=0.48$), no se cumplió el supuesto de normalidad en los tratamientos T1 y T2, lo que justificó el enfoque no paramétrico. Por su parte, la variable "Spam", que representa la duración del periodo de cosecha, mostró una correlación moderada con la densidad (0.36), aunque sin diferencias significativas entre tratamientos según Kruskal-Wallis. En conjunto, estos resultados indican que si bien la menor densidad retrasa el inicio de la cosecha, no prolonga necesariamente su duración, lo cual permite una planificación más predecible y eficiente del ciclo productivo.

Finalmente, el análisis estadístico implementado permitió confirmar la influencia significativa de la densidad de siembra sobre variables clave como el diámetro de la flor, el peso del tallo y el aprovechamiento del cultivo. La falta de cumplimiento de los supuestos de normalidad y homogeneidad de varianzas en varias variables validó el uso de pruebas no

paramétricas, mientras que la matriz de correlaciones de Kendall permitió identificar patrones direccionales relevantes. En conjunto, estos hallazgos confirman que herramientas como Python son altamente efectivas para abordar análisis agronómicos complejos, fortaleciendo la toma de decisiones con base en evidencia cuantitativa y estadísticamente robusta.

Recomendaciones

En el presente estudio se aplicaron diversas pruebas estadísticas para evaluar el efecto de la densidad de siembra sobre variables morfofisiológicas y productivas del cultivo de Brassicas. El análisis de normalidad mediante la prueba de Shapiro-Wilk y el análisis de homogeneidad de varianzas a través de la prueba de Levene, complementados con histogramas y gráficos Q-Q, revelaron que varias de las variables evaluadas no cumplen con los supuestos fundamentales necesarios para aplicar modelos paramétricos como el ANOVA. Esta situación, aunque no invalida los resultados, limita el tipo de análisis estadístico aplicable y sugiere la necesidad de realizar ajustes metodológicos en futuros estudios.

Es importante aclarar que el diseño experimental no presenta un desbalance entre tratamientos, ya que cada uno contó con un tamaño muestral cercano a las 30 observaciones, lo cual es adecuado desde una perspectiva básica del análisis estadístico. No obstante, este tamaño puede considerarse moderado en términos de potencia estadística. Para incrementar la sensibilidad de las pruebas y la precisión de las estimaciones, se recomienda aumentar el tamaño de muestra por tratamiento en futuros experimentos. Idealmente, una muestra de al menos 40 observaciones por tratamiento permitiría una mayor robustez frente a las desviaciones de normalidad o varianzas, especialmente cuando se planea el uso de pruebas paramétricas.

Respecto al tratamiento de datos no normales, se recomienda aplicar transformaciones que permitan cumplir los supuestos requeridos para utilizar análisis como el ANOVA. Para variables proporcionales como el porcentaje de tallos aprovechados, la transformación arcoseno de la raíz cuadrada es una opción clásica y efectiva para estabilizar la varianzas. En el caso de variables con sesgo a la derecha, como el peso por tallo o la longitud del tallo, se puede recurrir a transformaciones logarítmicas, de raíz cuadrada o Box-Cox, esta última especialmente útil para

encontrar la transformación óptima en función de la distribución observada. Tras aplicar dichas transformaciones, es necesario verificar nuevamente los supuestos mediante Shapiro-Wilk, Levene y los respectivos gráficos diagnósticos.

En situaciones donde las transformaciones no logren normalizar los datos ni homogeneizar las varianzas, se deben considerar alternativas metodológicas robustas. El uso de ANOVA de Welch representa una solución viable para comparar medias entre tratamientos sin asumir homocedasticidad. Asimismo, los modelos lineales generalizados (GLM), que permiten especificar distribuciones distintas a la normal (como Poisson, binomial o gamma), ofrecen una estructura flexible para ajustar los datos a modelos estadísticos más realistas.

Además, se recomienda realizar un análisis de potencia previo a la ejecución del experimento para definir el tamaño muestral necesario que permita detectar efectos con una probabilidad aceptable (generalmente superior al 80%). Esta práctica contribuiría a optimizar los recursos y asegurar resultados estadísticamente sólidos. También se sugiere el control y registro de covariables agronómicas o ambientales que puedan estar influyendo sobre las respuestas observadas; dichas variables pueden incorporarse en modelos ANCOVA para aumentar la precisión del análisis.

En síntesis, si bien el presente estudio estuvo correctamente estructurado en cuanto a la distribución de los tratamientos, se reconoce la necesidad de incrementar el tamaño de muestra, aplicar transformaciones cuando los datos no sean normales o no cumplan homogeneidad de varianzas, y considerar modelos estadísticos alternativos en caso de persistir dichas limitaciones. Estos ajustes permitirán reforzar la validez de las conclusiones y proporcionar un análisis estadístico más riguroso y representativo del efecto de la densidad de siembra sobre el cultivo de Brassicas.

Referencias

- Abril Rodríguez, J. S., Álvarez Herrera, J. G., & Jaime Guerrero, M. (2023). Evaluación de diferentes densidades de siembra en el cultivo de *Dianthus barbatus*, variedades Green Ball y Kiwi Mellow. *Ciencia e Ingeniería Neogranadina*, 33(2), 39–49.
- Asocolflores. (2023). *Informe de gestión y perspectivas del sector floricultor colombiano*.
- Box, G. E. P., Hunter, W. G., & Hunter, J. S. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2ª ed.). Wiley-Interscience.
- Brownlee, J. (2020). *Statistical modeling and machine learning in Python*. Machine Learning Mastery.
- Calero, A., Castillo, Y., Quintero, E., Pérez, Y., & Olivera, D. (2018). Efecto de cuatro densidades de siembra en el rendimiento agrícola del frijol común (*Phaseolus vulgaris* L.). *Revista de la Facultad de Ciencias*, 7(1), 88–100.
- Chavarría Párraga, J. E., & Vera Montenegro, L. O. (2010). Híbridos de girasol ornamental sometidos a diferentes densidades de siembra. *ESPAMCIENCIA*, 1(2), 60–64.
- Davidson-Pilon, C. (2015). *Bayesian methods for hackers*. Addison-Wesley.
- Departamento Administrativo Nacional de Estadística (DANE). (2022). *Cuentas nacionales anuales: Producto Interno Bruto por actividad económica*.
- Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241–252.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4ª ed.). SAGE Publications.
- Flor Verde. (2021). *Normas de buenas prácticas agrícolas en la floricultura colombiana*.
- Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, 24(3), 343–360.

- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486–489.
- Grupo GHT. (2023). *Portafolio de empresas: GHT Group*. <https://www.grupoght.com>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7^a ed.). Pearson.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
- Instituto Colombiano Agropecuario (ICA). (2013). *Resolución 00718 de 2013: Condiciones fitosanitarias para la exportación de productos agrícolas*.
- Instituto Colombiano Agropecuario (ICA). (2015). *Decreto 1843 de 2015: Directrices para la implementación de buenas prácticas fitosanitarias en el sector agrícola*.
- Instituto Colombiano Agropecuario (ICA). (2021). *Plan de sanidad vegetal para cultivos ornamentales*.
- Jones, E., Oliphant, T., & Peterson, P. (2001). *SciPy: Open source scientific tools for Python*. <https://www.scipy.org/>
- Kim, Y. R., Mascarini, L., Lorenzo, G. A., & González, M. N. (2018). Planting density and its effect on plant height and rosette quality in ornamental cabbages for cutting (*Brassica oleracea* L. var. *acephala*). *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo*, 50(2), 49–64.
- Kuehl, R. O. (2001). *Design of experiments: Statistical principles of research design and analysis*. Duxbury.

- McKinney, W. (2012). *Python for data analysis*. O'Reilly Media.
- McKinney, W. (2018). *Python for data analysis* (2^a ed.). O'Reilly Media.
- Millman, K. J., & Aivazis, M. (2011). Python for scientists and engineers. *Computing in Science & Engineering*, 13(2), 9–12.
- Ministerio de Agricultura y Desarrollo Rural (MADR). (1997). *Decreto 2191 de 1997: Organización y funcionamiento de entidades encargadas de la producción agrícola*.
- Ministerio de Agricultura y Desarrollo Rural (MADR). (2022). *Políticas y estrategias para el sector floricultor colombiano*.
- Ministerio de Trabajo. (2021). *Código Sustantivo del Trabajo y normas laborales en el sector agrícola*.
- Montgomery, D. C. (2012). *Design and analysis of experiments* (8^a ed.). Wiley.
- Mueller, A., & Massaron, L. (2016). *Introduction to machine learning with Python*. O'Reilly Media.
- Oliphant, T. E. (2006). *A guide to NumPy*. Trelgol Publishing.
- Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- ProColombia. (2023). *Oportunidades de exportación de flores colombianas*. Ministerio de Comercio, Industria y Turismo.
- Rainforest Alliance. (2020). *Normas para la certificación de sostenibilidad en la floricultura*.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.

- Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with Python*. Proceedings of the 9th Python in Science Conference, 57–61.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6^a ed.). Pearson.
- Tosi, A. (2020). *Data visualization with Python*. Packt Publishing.
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2), 22–30.
- VanderPlas, J. (2016). *Python data science handbook*. O'Reilly Media.
- Virtanen, P., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272.
- Waskom, M. (2017). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 2(11), 302.
- Yamane, T. (2011). *Statistics: An introductory analysis* (3^a ed.). Harper and Row.
- Yap, B. W., & Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141–2155.