R para Microbiología Industrial: Análisis de Datos y Diseño Experimental con un Enfoque Práctico

Fredy Ortiz Miguel Pérez Francisco León

2025-05-05

Tabla de contenidos

1 R para MI

Prefacio

"R para Microbiología Industrial: Análisis de Datos y Diseño Experimental con un Enfoque Práctico"

En el campo de la Microbiología Industrial y el diseño de experimentos, la integración de herramientas estadísticas constituye un desafío pedagógico fundamental que requiere estrategias innovadoras de enseñanza-aprendizaje, y como profesores de estas áreas de aprendizaje hemos identificado que los estudiantes experimentan dificultades significativas al establecer conexiones entre los conceptos estadísticos y los resultados experimentales microbiológicos.

En respuesta a esta problemática, surge la propuesta del libro " Aplicaciones del Software RStudio® en la Microbiología Industrial ", diseñado específicamente para articular las áreas de: Diseño de Experimentos y la Microbiología Industrial con ayuda de Rstudio®, utilizando a lo largo de contenido ejemplos concretos derivados de trabajos de grado y proyectos académicos desarrollados en la Universidad de Santander - UDES.

La obra integra además temas relacionados con Análisis Bibliométrico eInteligencia Artificial, reconociendo de este modo que la microbiología contemporánea demanda no solo competencias técnicas, sino adaptación de nuevas habilidades en una disciplina científica en constante evolución, contribuyendo de esta forma a la formación de profesionales capaces de afrontar los desafíos emergentes del campo de microbiológico industrial, tanto para el presente como su futuro profesional.

Fredy Alejandro Ortiz Meneses
Curso Microbiologia General y Microbiología II

Miguel Oswaldo Pérez Pulido
Curso Proyecto II – Microbiología Industrial
Maestría en Estadística Aplicada y Analítica de Datos

Francisco Javier León
Curso Proyecto I – Profesor de Microbiología Industrial
Maestría en Estadística Aplicada y Analítica de Datos

Lo que significa este libro

La presente obra constituye nuestra contribución a la formación integral de los Microbiólogos Industriales en su desarrollo como científicos. Esperamos que su contenido no solo fortalezca su experiencia académica, sino que además les provea de las competencias prácticas indispensables para afrontar los desafíos contemporáneos y futuros del ámbito profesional.

Para citar

Ortiz, F., Pérez, M., y León, F. (2025). *R para Microbiología Industrial: Análisis de Datos y Diseño Experimental con un Enfoque Práctico* Universidad de Santander. Vicerrectoría de Enseñanza.https://udesanalitica.github.io/r-para-mi/. https://doi.org/10.5281/zenodo.XXXXX

```
@techreport{PerezPulido2025,
  author       = {Ortiz, F., Pérez, M., y León, F.},
  title        = {R para Microbiología Industrial: Análisis de Datos y Diseño Experimental co
  institution  = {Universidad de Santander (UDES)},
  year         = {2025},
  note         = {Vicerrectoría de Enseñanza.
  url          = {https://udesanalitica.github.io/r-para-mi/}},
  type         = {e-book},
  doi          = {10.5281/zenodo.XXXXX},
  url          = {https://doi.org/10.5281/zenodo.XXXXX}
```

2 Autores

Tip

2.1 Fredy Alejandro Ortiz Meneses

Microbiólogo con énfasis en Alimentos, Especialista en Pedagogía y Didácticas Específicas, y Magíster en Fitopatología.

Tip

2.2 Miguel Oswaldo Pérez Pulido

Director de Analítica Académica. Licenciado en Matemáticas y Magíster en Estadística. Actualmente se desempeña como Director de Analítica Académica, adscrito a la Vicerrectoría de Enseñanza. Está vinculado a la Universidad de Santander (UDES) desde 2011, donde ha sido docente en programas de pregrado y posgrado de la Facultad de Ciencias Exactas, Naturales y Agropecuarias. Es investigador Junior reconocido por Minciencias en la convocatoria 894 de 2021 y miembro del grupo de investigación CIBAS.

Tip

2.3 Francisco Javier León

Bacteriólogo y laboratorista clínico, con formación avanzada como Magíster en Estadística Aplicada, Magíster en Ciencias Básicas Biomédicas y Especialista en Educación con Nuevas Tecnologías. Está vinculado a la Universidad de Santander (UDES) desde 2007, donde ha sido docente en la Facultad de Ciencias Exactas, Naturales y Agropecuarias. Actualmente, se desempeña como Coordinador de Analítica Académica, adscrito a la Vicerrectoría de Enseñanza. Es investigador Junior reconocido por Minciencias en la convocatoria 894 de 2021 y miembro del grupo de investigación CIBAS.

3 Agradecimientos

En primer lugar, queremos expresar nuestro más sincero agradecimiento a todos los estudiantes y profesores del programa de Microbiología que, a lo largo del tiempo, han compartido con nosotros sus inquietudes y retos al intentar conectar el análisis estadístico con la microbiología industrial.

Nuestro agradecimiento se extiende a los colegas académicos, especialmente a los profesores: Christian Andrey Chacín Zambrano y Daniel Adyro Martinez; y a los estudiantes graduados que generosamente compartieron sus experiencias y bases de datos, provenientes de importantes experimentos académicos. Sus aportes han sido fundamentales para dar vida a este manual y hacerlo relevante y aplicable a situaciones reales dentro del contexto de la microbiología industrial.

Asimismo, manifestamos gratitud a Robert Gentleman y Ross Ihaka, creadores del software R, así como a todos los colaboradores de la comunidad de R y RStudio®. Gracias a su compromiso y dedicación, estas herramientas se han mantenido accesibles para la comunidad científica.

A la Universidad de Santander (UDES) y a su Departamento de Desarrollo Profesoral, por la apertura de la Convocatoria Interna **Producción de Material Profesoral (2025)**, gracias a esta iniciativa, hemos encontrado un espacio de apoyo institucional que valora la producción material educativo de calidad, gracias a ello, nos sentimos motivados a seguir desarrollando herramientas que fortalezcan una enseñanza efectiva en los campos de la Microbiología Industrial y la Estadística Aplicada.

4 Introducción

En el ámbito de la microbiología industrial, donde los requerimientos analíticos varían según el tipo de experimento y los objetivos investigativos, R y RStudio® ofrecen una flexibilidad sobresaliente. La comunidad global de usuarios provee soporte constante y recursos actualizados, mientras que la amplia disponibilidad de paquetes especializados permite realizar análisis complejos con mayor precisión y eficiencia (Wickham & Grolemund, 2017). Esta combinación de potencia analítica, reproducibilidad y accesibilidad convierte a R en una herramienta idónea para el análisis de datos experimentales, el diseño de experimentos y la optimización de procesos biotecnológicos.

El presente libro está estructurado en tres partes principales.

- La Parte I aborda la instalación, configuración y manejo del entorno RStudio®, junto con un inventario de librerías esenciales para el análisis de datos en microbiología industrial.
- La Parte II desarrolla aplicaciones prácticas de R en el diseño experimental, con ejemplos reproducibles de diseños completamente al azar, en bloques y con mediciones repetidas en el tiempo.
- La Parte III introduce el uso de inteligencia artificial y simulación de datos, explorando cómo los modelos generativos y las herramientas computacionales pueden complementar la investigación microbiológica moderna.

De este modo, el libro ofrece una guía integral que combina fundamentos teóricos, práctica aplicada y perspectivas innovadoras, contribuyendo a fortalecer las competencias analíticas de los estudiantes y profesionales de la microbiología industrial.

5 Capítulo 1 Introducción al software R y la interfaz RStudio

El software **R** es un entorno de programación especializado en análisis estadístico, visualización de datos y modelado científico, ampliamente utilizado en la investigación y la industria. Su integración con **RStudio**, una interfaz de desarrollo amigable y versátil, facilita la escritura de código, la gestión de proyectos y la interpretación de resultados. A continuación, se describe el proceso de **instalación y configuración** del software R y de la interfaz **RStudio**, así como los pasos iniciales para familiarizarse con sus principales componentes y herramientas de trabajo.

5.1 Instalación y configuración

La instalación del software R y la interfaz RStudio (ahora llamado *Posit RStudio®*) es un proceso sencillo que puede completarse en unos pocos pasos; primero, se debe descargar e instalar R desde el sitio web oficial del Proyecto R (https://www.r-project.org); una vez instalado R, se puede proceder a descargar e instalar RStudio® desde su sitio web (https://posit.co/download/rstudio-desktop/) (Figura 1).

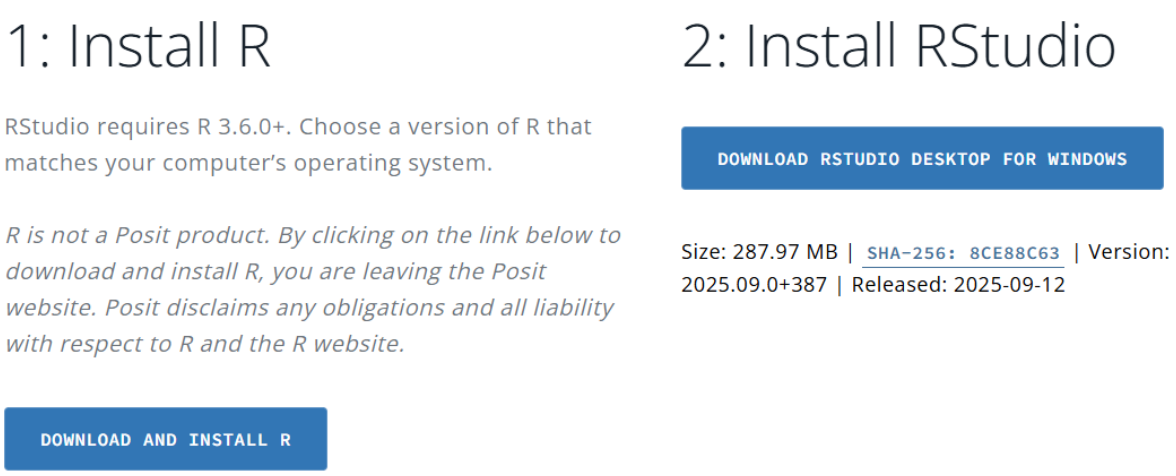


Figura 5.1: Figura 1.

Ambos programas están disponibles para múltiples sistemas operativos, incluyendo Windows, macOS y Linux (Figura 2).

Figura 5.2: Figura 2.

Del mismo modo se deben descarga de diferentes directorios llamadas CRAN (Comprehensive R Archive Network o Red integral de archivo R) (Figura 3).

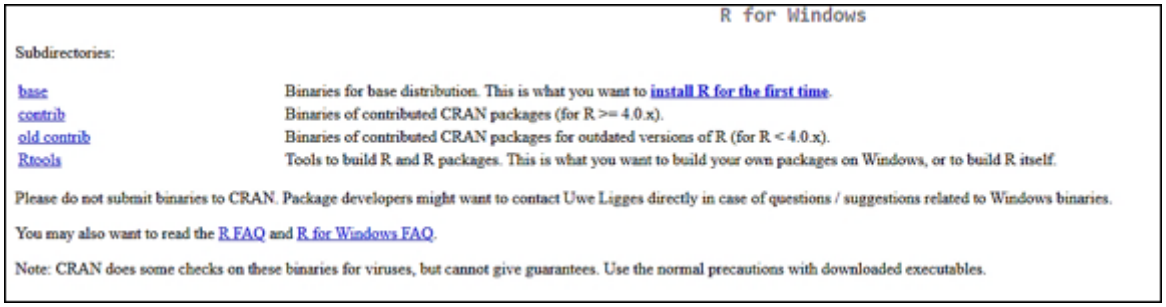


Figura 5.3: Figura 3.

Finalmente descargar la última versión de R para Windows (Figura 4).

Figura 5.4: Figura 4.

Una vez instalados R y RStudio®, es importante familiarizarse con la interfaz de RStudio® (Figura 5); esta interfaz está dividida en varias secciones, incluyendo: (i) el editor de código o Script, el cual permite escribir y editar las instrucciones o Scripts; (ii) la consola: se utiliza para ejecutar comandos interactivos; (iii) el entorno de trabajo: muestra los objetos y datos cargados en la sesión actual y las (iv) pestañas de archivos y gráficos: permiten gestionar archivos y visualizar gráficos generados por R (R Core Team, 2021).

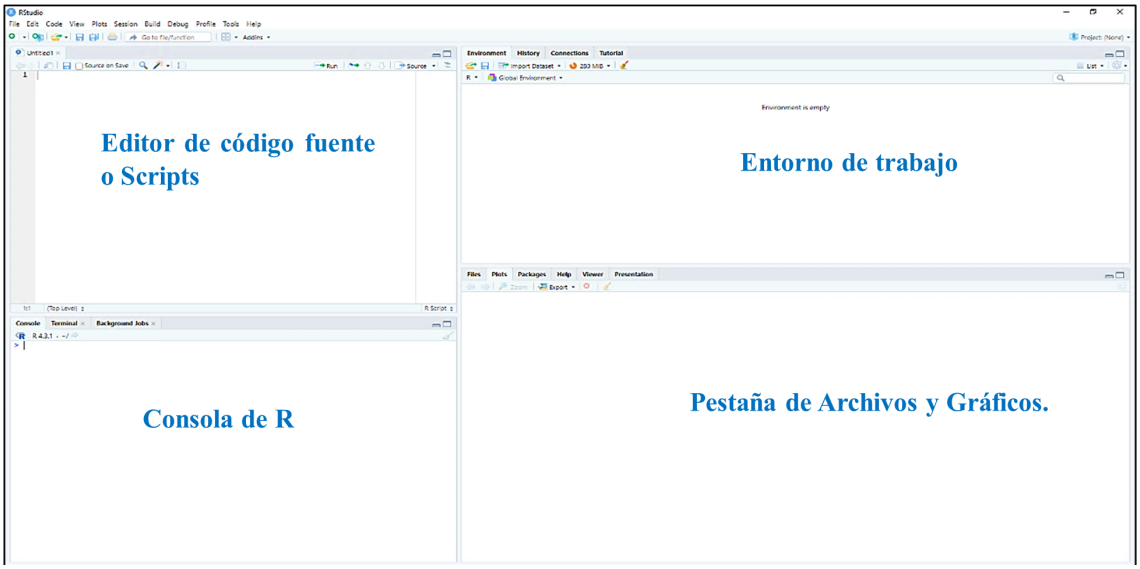


Figura 5.5: Figura 5

Además de la interfaz básica, RStudio® permite la instalación y gestión de paquetes adicionales que amplían sus funcionalidades; para instalar un paquete, se puede utilizar la función install.packages ("nombre_del_paquete") en la consola de RStudio®; una vez instalado, el paquete se puede cargar en la sesión actual utilizando la función library (nombre_del_paquete).

> **!** Importante
>
> Mantener R y RStudio® actualizados es clave para aprovechar las últimas mejoras, nuevas funcionalidades y correcciones de errores. Ambos programas notifican automáticamente cuando hay versiones más recientes disponibles, por lo que se recomienda estar atento a estos avisos y actualizar oportunamente.

Para actualizar R, se debe descargar e instalar la nueva versión desde el sitio web del Proyecto R; para actualizar RStudio®, se puede utilizar la opción de actualización en el menú de ayuda de RStudio®; mantener el software actualizado garantiza un rendimiento óptimo y acceso a las últimas funcionalidades (R Core Team, 2021) (R Core Team, 2023; RStudio Team, 2023).

5.2 Paquetes Esenciales para el análisis de datos

De **R** se destaca su gran variedad de paquetes especializados que amplían sus capacidades analíticas y gráficas. En el contexto de la **microbiología industrial**, estas librerías permiten gestionar, transformar y visualizar datos experimentales con precisión, favoreciendo la interpretación de resultados y la toma de decisiones basadas en evidencia. Es por ello que se presenta un inventario de los **paquetes más utilizados en el análisis de datos microbiológicos**.

5.3 Inventario de Librerías y Paquetes de R aplicados para el análisis de datos en Microbiología Industrial.

El ecosistema de librerías y paquetes de R constituye una herramienta fundamental para el análisis de datos en microbiología industrial, proporcionando soluciones específicas para cada etapa del proceso investigativo, y en este contexto, las librerías básicas como:

- ***readxl***, desarrollada por (Wickham & Bryan, 2015), facilita la importación de datos desde hojas de cálculo Excel®, donde tradicionalmente los microbiólogos registran sus resultados experimentales.
- ***car*** (Companion to Applied Regression), creada por (Fox & Weisberg, 2019), ofrece herramientas esenciales para la verificación de supuestos estadísticos mediante gráficos QQ-plot, permitiendo evaluar la normalidad de los datos antes de aplicar pruebas

paramétricas en experimentos de optimización de medios de cultivo y comparación de cepas microbianas.

La revolución en el análisis de datos microbiológicos se materializa principalmente a través del libreria ***tidyverse***, desarrollado por (Wickham et al., 2019), que integra múltiples librerías bajo una lógica común de programación, Este conjunto incluye los siguientes paquetes:

- **ggplot2** , para visualización de datos.
- **dplyr** , para manipulación de datos.
- **tidyr** , para ordenar datos.
- **readr** , para importar datos.
- **purrr** , para programación funcional.
- **tibble** , para tibbles, una reinvención moderna de los marcos de datos.
- **stringr** , para cadenas.
- **forcats** , para factores.
- **lubridate** , para fecha/hora.

Los analísis se enriquecen, considerablemente con librerías especializadas que abordan necesidades específicas de la investigación microbiológica industrial, y tal es el caso de:

- ***gridExtra***, desarrollada por (Auguie, 2017), la cual facilita la organización de múltiples gráficos en una sola visualización, permitiendo comparaciones efectivas entre diferentes condiciones experimentales.
- ***lsr*** (Learning Statistics with R), creada por (Navarro, 2015) proporciona funciones accesibles para análisis estadísticos fundamentales como pruebas t, ANOVA y cálculos de tamaño del efecto;
- **Bibliometrix**, desarrollado por (Aria & Cuccurullo, 2017) permite realizar análisis bibliométrico de publicaciones científicas, identificando tendencias emergentes y redes de colaboración que orientan nuevas investigaciones.

Las aplicaciones especializadas en análisis multivariado y modelado avanzado complementan este inventario tecnológico, y es donde converge:

- ***vegan***, desarrollada por (Oksanen et al., 2020), la cual proporciona proporciona herramientas para análisis de diversidad ecológica mediante técnicas como PCA (Análisis de Componentes Principales), NMDS (Escalamiento Multidimensional No Métric) permitiendo visualizar relaciones complejas entre comunidades microbianas y variables ambientales en procesos industriales.

- ***nlme*** desarrollado por (Pinheiro et al., 2025) ofrece capacidades para modelar datos longitudinales con estructura jerárquica, típicos de estudios de cinética microbiana.
- ***Agricolae*** facilita el diseño experimental(Mendiburu, 2020).
- ***Shiny*** permite desarrollar aplicaciones web interactivas para visualización dinámica de resultados, mejorando la colaboración y transparencia en la investigación microbiológica industrial(Chang et al., 2021).

5.4 Paquetes del software R para Microbiología Industrial

El uso de R como herramienta de análisis estadístico en la microbiología industrial ha experimentado un crecimiento exponencial en la última década, según (Mohammadi et al., 2019) R proporciona una plataforma versátil que permite analizar datos complejos derivados de experimentos microbiológicos, facilitando la identificación de patrones de crecimiento microbiano, optimización de condiciones de cultivo y evaluación de la producción de metabolitos secundarios, lo que resulta crucial para el desarrollo y mejora de procesos biotecnológicos en entornos industriales.

- El paquete *phyloseq* (McMurdie & Holmes, 2013), se emplea en el análisis de datos de secuenciación en estudios de comunidades microbianas, permitiendo la integración de información taxonómica, filogenética y de abundancia en un solo entorno analítico; este avance ha sido fundamental para comprender la dinámica de **poblaciones microbianas** en procesos industriales como: el tratamiento de aguas residuales, la producción de biocombustibles y la fermentación alimentaria.
- El paquete microbiome, descrito por (Lahti & Shetty, 2017), proporciona herramientas especializadas para el análisis de **datos metagenómicos**, facilitando la caracterización de comunidades microbianas y sus funciones metabólicas en entornos industriales, lo que resulta esencial para la optimización de bioprocesos y el control de calidad en la industria alimentaria.

El **diseño experimental** en microbiología industrial se ha beneficiado significativamente de la aplicabilidad de R, permitiendo planificar y analizar experimentos de manera más rigurosa y eficiente.

- El paquete *agricolae*, desarrollado por de Mendiburu (2021) (Zhou et al., 2012) es utilizado para la implementación de diseños experimentales complejos como: bloques aleatorizados y diseños factoriales entre otros, al tiempo que frecuentemente son utilizados en estudios de optimización de medios de cultivo, condiciones de fermentación y producción de enzimas microbianas.

- Complementariamente, (Ritz & Streibig, 2005) presentaron el paquete *drc* (Dose-Response Curves), que ha facilitado el análisis de **Curvas dosis-respuesta** en estudios de inhibición microbiana, pruebas de susceptibilidad a antimicrobianos y evaluación de compuestos bioactivos producidos por microorganismos, proporcionando herramientas estadísticas robustas para cuantificar y modelar respuestas biológicas a diferentes tratamientos, lo cual es fundamental en el desarrollo de nuevos productos biotecnológicos.
- El paquete *ggplot2* desarrollado por (Wickham, 2016), el cual ha permitido la creación de gráficos altamente informativos que facilitan la interpretación de resultados experimentales; en particular, la representación gráfica de cinéticas de crecimiento microbiano, producción de metabolitos y análisis multivariantes se ha vuelto más accesible e intuitiva para investigadores en el campo.
- De manera similar el paquete *ggtree*, creado por (Yu et al., 2017), ha revolucionado la visualización de datos filogenéticos en estudios de diversidad microbiana industrial, permitiendo representar relaciones evolutivas entre microorganismos de interés biotecnológico y correlacionarlas con características fenotípicas relevantes para procesos industriales, lo que facilita la selección de cepas microbianas con potencial biotecnológico.

Expandir para aprender son el analis de datos ómicos

El análisis de datos ómicos en microbiología industrial se ha visto significativamente potenciado gracias al aporte de (Love et al., 2014) quienes introdujeron *DESeq2*, un paquete que ha transformado el análisis de datos de RNA-seq en estudios transcriptómicos de microorganismos industriales, permitiendo identificar genes diferencialmente expresados bajo diversas condiciones de cultivo o modificaciones genéticas; lo que contribuye a la mejora de cepas microbianas industriales y a optimizar rutas metabólicas de interés comercial; paralelamente (Rohart et al., 2017) desarrollaron el paquete mixOmics, el cual facilita la integración de múltiples conjuntos de datos ómicos, como:

(i) transcriptómica,
(ii) proteómica y
(iii) metabolómica,

Proporcionando una visión holística de los sistemas microbianos en contextos industriales, lo que permite desentrañar complejas redes regulatorias y metabólicas que subyacen a procesos biotecnológicos importantes como de compuestos bioactivos.

6 Capítulo 2 Análisis bibliométrico para la gestión de un diseño experimentos

El **Capítulo 2** aborda el análisis bibliométrico como una herramienta esencial para la **gestión del conocimiento en el diseño experimental**, permitiendo identificar tendencias, autores, revistas y enfoques metodológicos relevantes dentro del campo de la microbiología industrial. La inclusión de este capítulo tiene como propósito **fortalecer la fundamentación teórica y metodológica** del lector, proporcionándole una visión panorámica de la producción científica relacionada con el tema. A través del uso de técnicas de minería de datos bibliográficos y del software *R*, se ejemplifica cómo la bibliometría contribuye a **orientar la planificación y optimización de los diseños experimentales**, facilitando la toma de decisiones basadas en evidencia científica actual. De este modo, el capítulo no solo enriquece el marco conceptual del libro, sino que también promueve una práctica investigativa más informada, sistemática y alineada con las tendencias globales en investigación aplicada.

6.1 Etapas del análisis bibliométrico

El análisis bibliométrico es un proceso estructurado que permite examinar de manera sistemática la producción científica sobre un tema determinado. Para garantizar resultados rigurosos y reproducibles, este procedimiento se desarrolla en varias etapas que van desde la **definición del tema y la búsqueda en bases de datos**, hasta la **depuración, análisis e interpretación de los resultados** mediante herramientas especializadas como *Bibliometrix* y su interfaz *Biblioshiny*.

A continuación, se describen las etapas del proceso:

6.1.1 1. Definición del tema y palabras clave

La primera etapa consiste en **delimitar el tema de estudio** y seleccionar las **palabras clave** que representen el objeto de investigación. Por ejemplo, en este ejercicio se emplearon los datos del trabajo de grado (sin publicar) *"Evaluación del crecimiento de Cordyceps militaris en diferentes sustratos vegetales"* (Chala, 2025).

Algunas palabras clave empleadas fueron:

- *Cultivation*: proceso de cultivar *Cordyceps militaris* en condiciones controladas para estudiar su crecimiento.
- *Mycelial growth*: crecimiento del micelio, parte vegetativa del hongo.
- *Substrate optimization*: mejora de los sustratos de cultivo para maximizar el crecimiento y la producción de metabolitos.
- *Bioactive compounds*: compuestos bioactivos como la cordicepina, con propiedades medicinales.
- *Fermentation conditions*: condiciones de fermentación (temperatura, pH, nutrientes) que influyen en el crecimiento del hongo.

Las combinaciones de estas palabras se construyen utilizando **operadores booleanos** (AND, OR, NOT) para lograr ecuaciones de búsqueda precisas.

Por ejemplo:

```
"Cordyceps militaris" AND ("substrate optimization" OR "culture
medium") AND "growth"
```

6.1.2 2. Búsqueda y descarga de información en bases de datos

Una vez definidas las palabras clave, se realiza la **búsqueda sistemática** en bases de datos compatibles con *Bibliometrix*, como: **Web of Science** , **Scopus**, **OpenAlex**, **Dimensions**, **The Lens**; **PubMed**y **Cochrane Library**

Durante esta etapa se deben aplicar filtros de búsqueda: **Periodo de tiempo, Tipo de documento** (artículo, revisión, conferencia), **Área temática.**

Luego, los **metadatos** deben descargarse en formato **CSV, BibTeX o RIS** según lo admita la base. Por ejemplo, en Scopus puede exportarse como *CSV (UTF-8)* con la opción "Full Record". Se recomienda nombrar los archivos de forma clara (p. ej. `Cordyceps_Scopus_2025.csv`).

En la Tabla 1 se presenta un ejemplo de resultados obtenidos tras aplicar distintas ecuaciones de búsqueda relacionadas con el tema.

Tabla 6.1: Tabla 1. Salida de resultados para cada uno de los operadores booleanos, introducidos dentro de la plataforma de Scopus®, y que están con el tema de investigación de *Cordyceps militaris.*

Ecuación de búsqueda	Documentos encontrados
"Cordyceps militaris" AND ("metabolite" OR "growth conditions"	225

Ecuación de búsqueda	Documentos encontrados
“Cordyceps militaris” AND (“cordycepin”) AND “growth”	191
“Cordyceps militaris” AND “medium” AND “growth”	99
“Cordyceps militaris” AND (“substrate optimization” OR “culture medium”) AND “growth”	40

6.1.3 3. Importación y análisis en Bibliometrix / Biblioshiny

Con los archivos descargados, se procede a su análisis mediante *Bibliometrix*, un paquete de R que facilita la recopilación, análisis y visualización de información científica de forma integral (Aria & Cuccurullo, 2017). La interfaz *Biblioshiny* permite ejecutar estos análisis de manera interactiva y sin necesidad de programación. **Bibliometrix** (a través de su interfaz **Biblioshiny**) está estructurado en **8 módulos principales** o menús, y dentro de cada uno hay **submódulos o indicadores específicos**.

Tabla 6.2: Módulos principales de Bibliometrix / Biblioshiny

Nº	Módulo	Función general	Submódulos / indicadores
1	**Data**	Importar/cargar bases bibliográficas (Scopus, WoS, PubMed, etc.).	Import or Load; Merge Datasets
2	**Filters**	Filtrar el corpus por año, tipo de documento, autores, países, palabras clave.	Time Span; Authors; Countries
3	**Overview**	Panorama general con indicadores descriptivos.	Main Information; Annual Scientific Production; Average Citations per Year; Three-Field Plot
4	**Sources**	Análisis de revistas/fuentes.	Most Relevant Sources; Bradford’s Law
5	**Authors**	Productividad, impacto y colaboración de autores.	Authors’ Production Over Time; Most Cited Authors; Collaboration Network

Nº	Módulo	Función general	Submódulos / indicadores
6	**Documents**	Documentos más citados y patrones de citación.	Most Cited Documents; Reference Spectroscopy
7	**Clustering (Conceptual Structure)**	Estructura temática/conceptual del campo.	Co-occurrence Network; Thematic Map; Factorial Analysis
8	**Social Structure**	Redes de colaboración entre autores, instituciones y países.	Collaboration Network; Country Scientific Production; Collaboration World Map

6.1.4 4. Interpretación y comunicación de resultados

Los resultados obtenidos se interpretan según el contexto de estudio. En el caso de la **microbiología industrial**, el uso de *Bibliometrix* permite identificar líneas emergentes de investigación, autores influyentes, colaboraciones internacionales y vacíos en la literatura.

A través de las visualizaciones interactivas de *Biblioshiny*, es posible construir **mapas de conocimiento**, **redes de colaboración** y **tendencias temáticas**, que facilitan la toma de decisiones basadas en evidencia científica.
De esta forma, el análisis bibliométrico se consolida como una herramienta que fortalece la **planificación de proyectos**, la **vinculación académica** y la **comprensión del panorama científico actual** en el campo de la microbiología industrial.

6.2 El paquete Bibliometrix

Desarrollado en R, constituye una herramienta clave para el análisis bibliométrico en distintas áreas del conocimiento, entre ellas la **microbiología industrial** y el **diseño experimental**. Su enfoque de código abierto permite recopilar, analizar y visualizar información científica de manera integral, ofreciendo una visión clara sobre las principales tendencias y la evolución de la investigación en cada campo (Aria & Cuccurullo, 2017).

En la **microbiología industrial**, **Bibliometrix** se ha convertido en un apoyo fundamental para reconocer **líneas emergentes de investigación**, **colaboraciones internacionales** y **autores influyentes** que marcan el desarrollo del área (Aria & Cuccurullo, 2017). A través de su interfaz visual **Biblioshiny**, los análisis complejos se vuelven accesibles incluso para quienes no tienen experiencia en programación. Esta accesibilidad favorece la creación de **mapas de conocimiento**, **redes de colaboración** y **agrupamientos temáticos** que

ayudan a identificar oportunidades de trabajo conjunto, vacíos en la literatura o la evolución de determinadas técnicas experimentales.

El uso de Bibliometrix y Biblioshiny permite una comprensión argumentada del panorama científico, fomentando decisiones de investigación basadas en evidencia y fortaleciendo la planificación de proyectos dentro de la microbiología industrial.

6.2.1 Instalación de bibliometrix

Para iniciar el análisis bibliométrico, se procede a ingresar al programa RStudio®, al tiempo que se realiza la instalación de los paquetes: *bibliometrix* y *bibliometrixData*, desde la pestaña de Archivos y Gráficos en la sección de Packages (Figura 8).

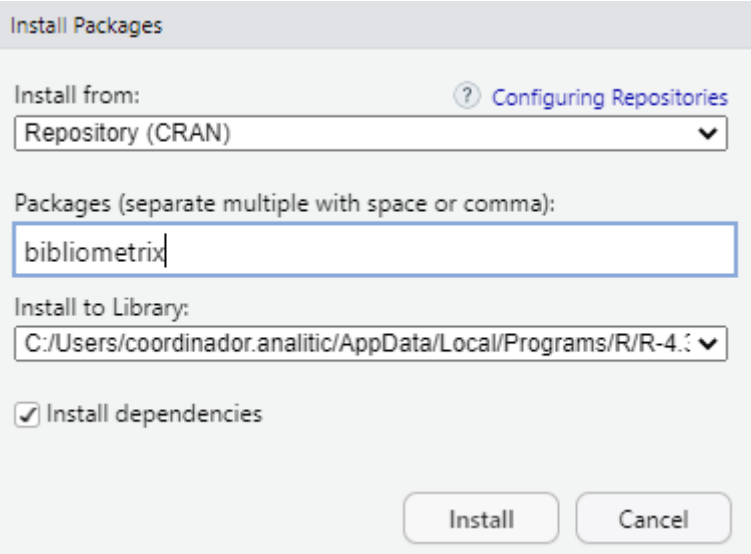


Figura 6.1: Figura 8.

Posteriormente, desde la Consola de RStudio se digitan y ejecutan los comandos:

```
library(bibliometrix)
biblioshiny()
```

Tambien se puede habilitar manualmente en la pestaña de Archivos y Graficos de la interfaz de RStudio, pestaña de Packages (figura9) seleccionado las librerias a usar.

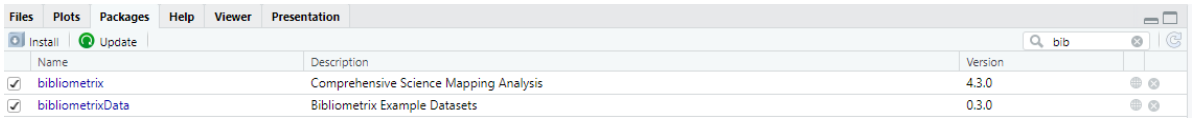


Figura 6.2: Figura 9.

Se abrirá el servidor de Bibliometrix (Figura 9) en el navegador web (Chrome, Mozilla, Edge, entre otros). Es importante aclarar que la interfaz de Bibliometrix únicamente pude ser ejecutada desde RStudio®.

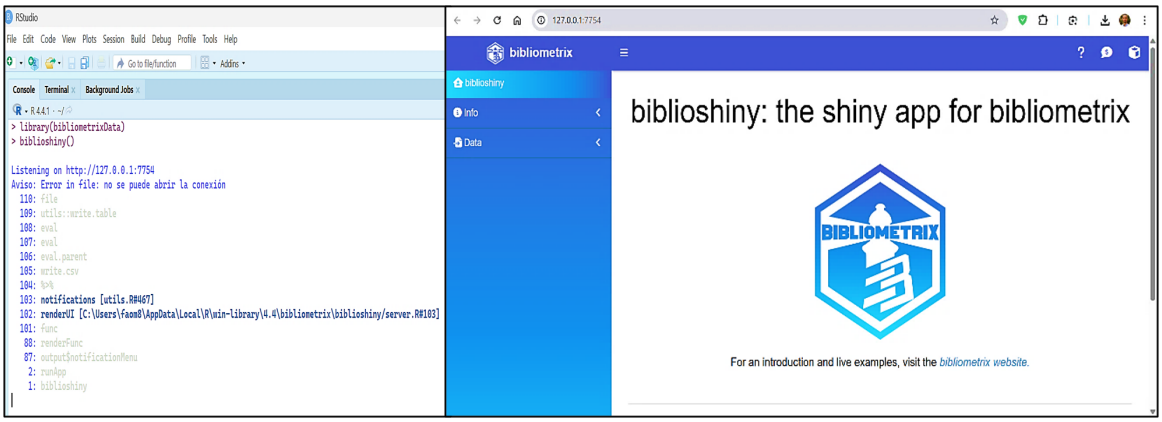


Figura 6.3: Figura 9.

En la parte izquierda se despliega el **menú de biblioshiny**, y se procede con la importación en "***Import or Load***" se carga el archivo CSV (Aria & Cuccurullo, 2017), al tiempo que se seleccionan las casillas: ***Import raw file(s)***, la procedencia de la base de datos consultada (*Scopus,* en nuestro caso) , junto la opción: ***Surname and Initials***, y finalmente damos click en el botón de ***Start*** (Figura 10).

Figura 6.4: Figura 10.

Después se despliega una nueva ventana que muestra el estado de los componentes de los metadatos importados desde el archivo CSV (Figura 11), allí se muestra una tabla que resume la completitud de los metadatos (para nuestro ejemplo: 191 documentos de Scopus),.

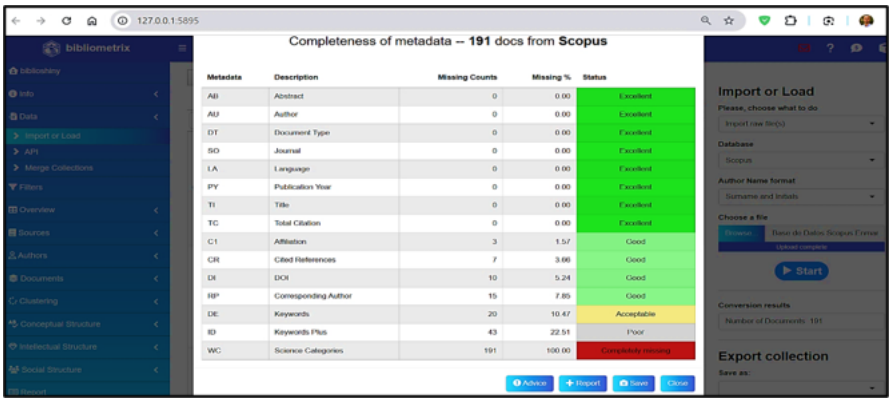


Figura 6.5: Figura 11.

Así mismo Bibliometrix evalúa cada uno de los diferentes campos de metadatos (Abstract, afiliación, autor, tipo de documento, etc.) utilizando *junto a* diferentes criterios de clasificación como son: “Excelente”, “Bueno”, “Aceptable” “Pobre” y “Completamente perdido” (Figura 11), y que al ser cerrado “Close” muestra una tabla dinámica con todos los datos importados incluyendo el DOI, desde el cual se puede acceder y leer directamente el articulo científico (Figura 12).

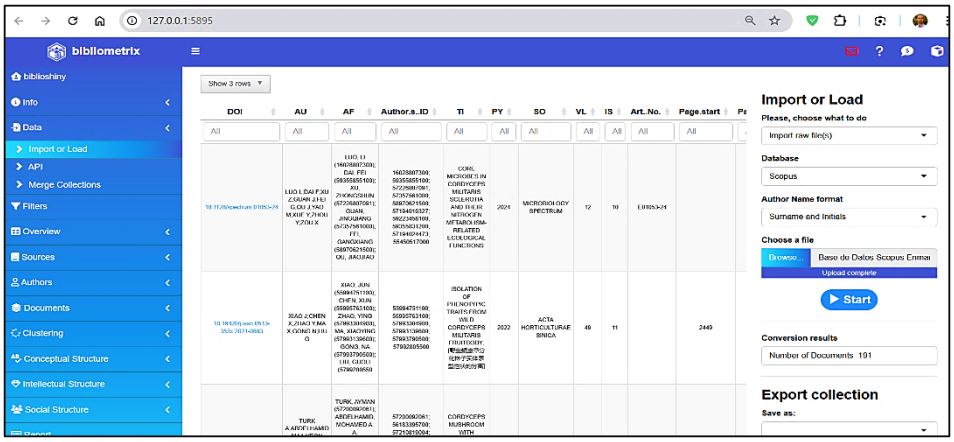


Figura 6.6: Figura 12

6.2.2 Estructura

6.2.2.1 Modulo *Overview*

El indicador ***Main information*** muestra los valores de los metadatos trabajados, para este ejemplo se tiene un periodo de análisis de1951 a 2025 con un total de 135 fuentes y 191 documentos, la tasa de crecimiento anual es de 0,94%, **se identificaron 903 autores**, sin registros un solo autor; la coautoría internacional alcanza un valor de 15,18% y el promedio de **coautores** por documento es de 6,19. Se encontraron 543 palabras clave de autor y 7474 referencias, la edad promedio de los documentos es de 7,34 años y la cantidad **media de citas por documento** es de 25,69 (Figura 13).

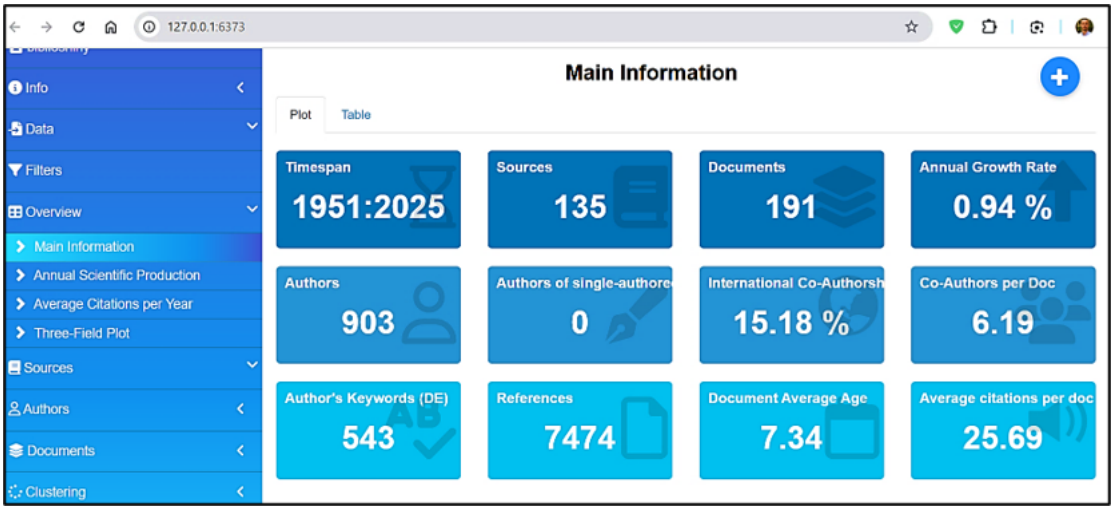


Figura 6.7: Figura 13

Para la sección ***Average Citations Per Year*** (Traducido: Citas Promedio por año) muestra la evolución de publicaciones sobre *Cordyceps militaris* entre 1951 y 2025 (Figura 14). Entre **1951 y 2000**, la producción científica fue mínima, con casi nula variación. A partir de **2000**, comienza un crecimiento leve y sostenido, que se **acelera notablemente después de 2010**, alcanzando su punto máximo entre **2020 y 2024**, con más de **20 artículos anuales**. En **2025**, se observa una **caída abrupta**, posiblemente atribuida a datos incompletos o publicaciones aún en proceso. En síntesis, la tendencia general evidencia un **crecimiento exponencial de la investigación** sobre *Cordyceps militaris* en las dos últimas décadas, reflejando su creciente relevancia científica y biotecnológica.

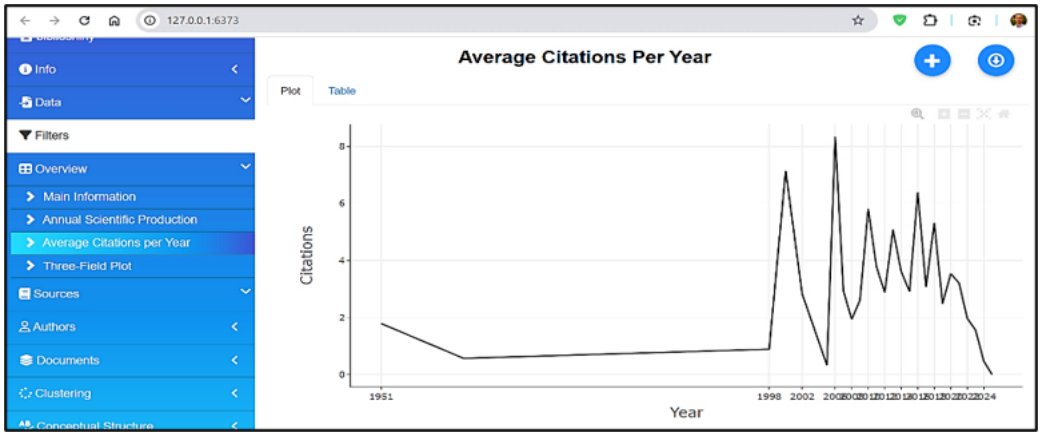


Figura 6.8: Figura 14

La Figura 15 **"Three-Field Plot"** muestra la relación entre **países (AU_CO)**, **autores**

(AU) y **descriptores temáticos (DE)** en la investigación sobre *Cordyceps militaris*. **China** lidera ampliamente la producción científica, seguida por **Tailandia**, **Reino Unido**, **Singapur** y **Estados Unidos**, evidenciando una fuerte concentración asiática en el tema. Los autores más productivos son **Li X**, **Li Y**, **Vongsangnak W** y **Zhang J**, quienes forman redes de colaboración relevantes dentro del campo. En cuanto a los **descriptores**, predominan términos como *Cordyceps militaris*, *biotecnología del hongo*, *compuestos bioactivos* y *análisis transcriptómico*, lo que refleja el enfoque de la investigación en los componentes biológicos y aplicaciones biotecnológicas del hongo. El gráfico confirma el **liderazgo científico de Asia** en el estudio de *Cordyceps militaris* y la **concentración temática en la exploración de compuestos bioactivos y su potencial en salud y biotecnología.*

quarto

6.2.2.2 Modulo *Sources*

Para el menú de *Sources* en la sección ***Most Relevant Sources*** (Traducido: Fuentes más Relevantes) la producción científica está liderada por: International Journal of Medicinal Mushrooms con 15 artículos publicados, seguido de Mycosystema con 8, le sigue: Applied Microbiology and Biotechnology con 5. Dichas revistas destacan por su enfoque en microbiología, biotecnología y farmacología, áreas clave dentro del estudio abordado. **La distribución sugiere que la investigación en este campo se encuentra bien representada en revistas especializadas**,las demás revistas como: Biology, Bioresource Technolgy, y Nutrients, que cuentan con 3 artículos cada una (Figura 16).

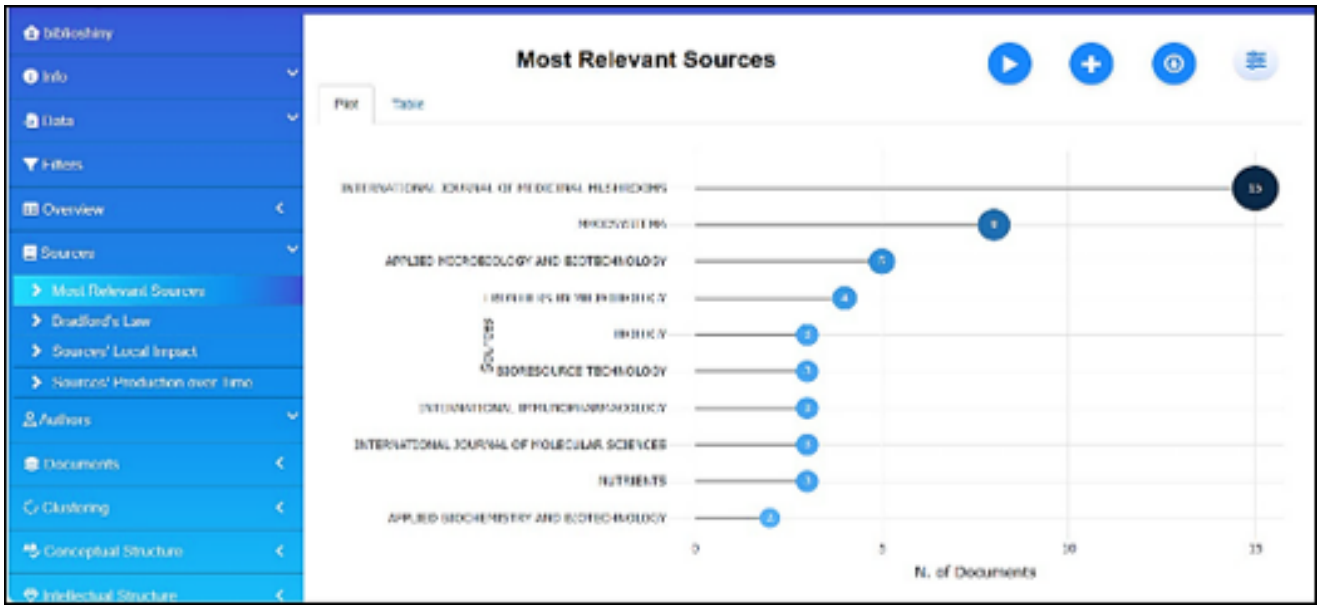


Figura 6.9: Figura 16

En la sección de **Bradford's Law** (Traducido: La Ley de Bradford) y continuando con nuestro ejemplo didáctico de *Cordyceps mil*itaris ,se observa que: "International Journal of Medicinal Mushrooms", junto con "Mycosystema" y "Applied Microbiology and Biotechnology", conforman el núcleo de fuentes indexadas más relevantes, aportando el mayor número de publicaciones, estas tres revistas están dentro del área sombreada, lo que **confirma su papel central en la diseminación del conocimiento** sobre *C. militari*s y compuestos bioactivoss. A medida que se avanza hacia la derecha del gráfico, el número de artículos por revista disminuye, lo que representa publicaciones de interés más disperso (Figura 17).

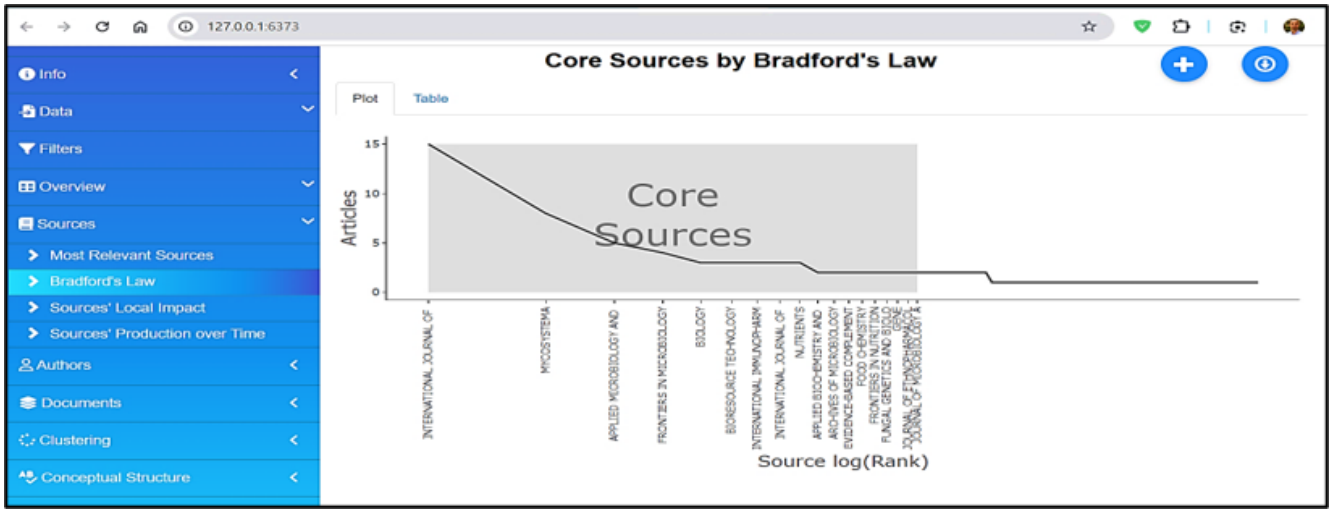


Figura 6.10: Figura 17

6.2.2.3 Modulo *Authors*

En la sección **"Authors' Production over Time"** (Producción de los autores a lo largo del tiempo) se evidencia que los investigadores **Li X., Li Y. y Wang Y .** han mantenido una **producción científica constante** en los últimos años, alcanzando **picos destacados en 2020 y 2022** (Figura 18), lo que demuestra su papel central en el estudio de *Cordyceps militaris*. La gráfica muestra una **mayor concentración de publicaciones entre 2019 y 2024**, lo que refleja un **crecimiento sostenido y reciente de la investigación** en este campo. Otros autores, como **Vongsangnak W.**, **Zhang J.**, **Laoteng K.** y **Thanwisai R.**, presentan una participación **más intermitente**, aunque continúan contribuyendo activamente en colaboraciones científicas internacionales. En conjunto, la visualización confirma una **expansión continua de la productividad académica**, impulsada por investigadores consolidados y por el aumento del interés global en las aplicaciones biotecnológicas y farmacológicas del hongo *Cordyceps militaris* .

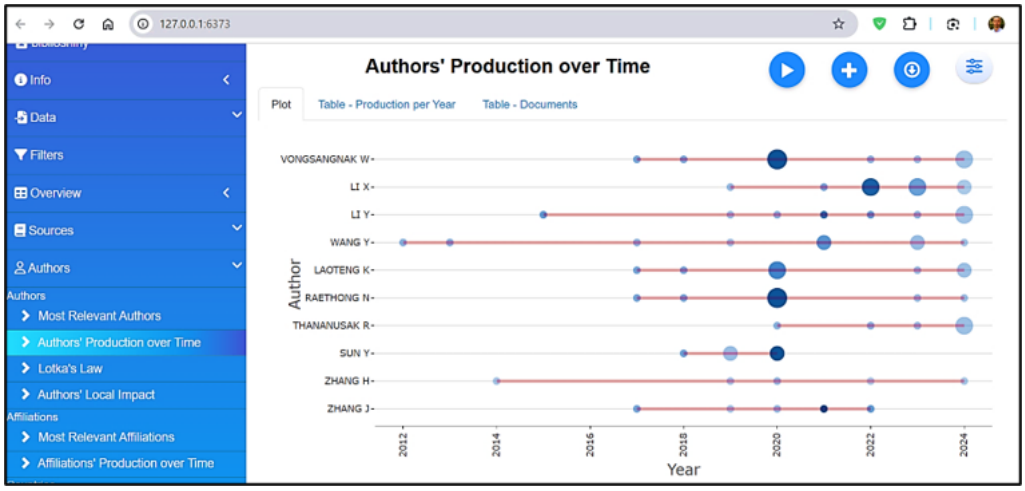


Figura 6.11: Figura 18

Para el menú de *Authors* específicamente en: ***Countries' Scientific Production*** (traducido: Producción científica de los países), el mapa muestra la distribución geográfica de la producción científica (Figura 19). China es el país con mayor producción científica (azul oscuro); otros países con destacada producción científica entre los que se incluyen: Estados Unidos, Corea del Sur, Tailandia, Japón, India y varios países europeos y asiáticos (azul celeste). **Algunos países no presentan producción registrada y aparecen coloreados en gris.** En la tabla se observa que China lidera con 702 publicaciones, seguida por Corea del Sur con 212, Tailandia con 83, Japón con 50 e India con 39. Otros países con menor producción incluyen Estados Unidos con 9, Reino Unido con 8, Alemania con 7, Italia con 6 y Colombia con 5 .

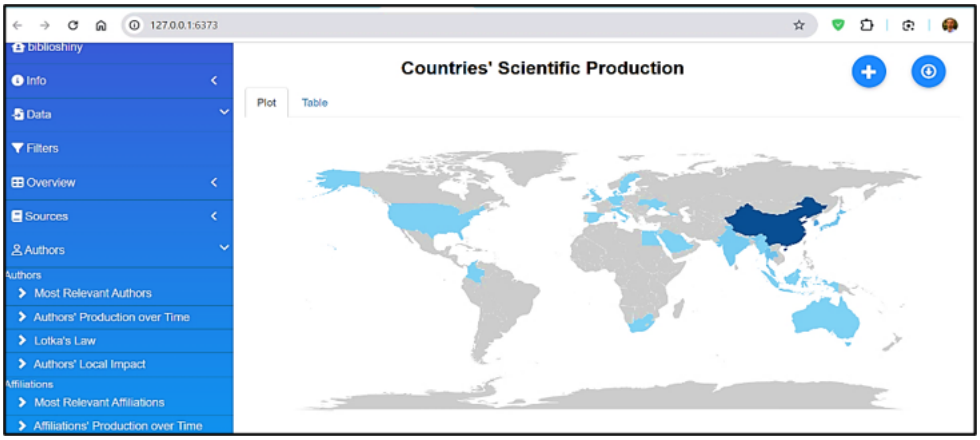


Figura 6.12: Figura 19

6.2.2.4 Modulo *Document*

En la sección **"Most Frequent Words"** (Palabras más frecuentes) del modulo **Documents**, la visualización identifica los términos que aparecen con mayor recurrencia en la literatura científica sobre *Cordyceps militaris*. Las palabras **"Cordyceps"** y **"Cordycepin"** destacan con **196** y **187** menciones respectivamente, reflejando su relevancia central en los estudios del área. Otros términos con alta frecuencia son **"article"** (109), **"Cordyceps militaris"** (108), **"nonhuman"** (92), **"metabolism"** (79), **"deoxyadenosines"** (77), **"controlled study"** (76), **"deoxyadenosine derivative"** (65) y **"adenosine"** (64). En conjunto, esta distribución de palabras clave evidencia que la investigación reciente se concentra en los **aspectos bioquímicos y farmacológicos** del hongo, especialmente en torno a sus compuestos activos y su aplicación en estudios experimentales (Figura 19).

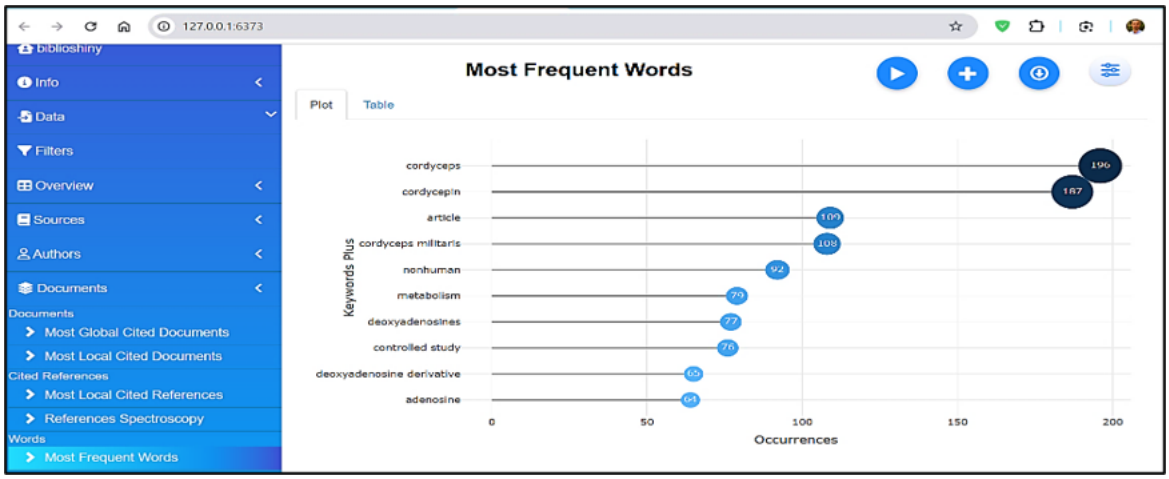


Figura 6.13: Figura 20

En la sección **"Reference Spectroscopy"** (Espectroscopía de referencias) del modulo **Documents**, la visualización muestra la evolución temporal de las **referencias citadas** en estudios de espectroscopía. Antes de **1990**, las citas registradas son casi inexistentes, indicando una actividad investigativa limitada. A partir de **1995**, se observa una **tendencia ascendente constante**, que se intensifica de forma notable hacia **2005**, evidenciando un crecimiento sostenido en la producción científica y en el interés por la temática. El **pico máximo** se alcanza entre **2015 y 2020**, con más de **400 referencias citadas por año**, lo que refleja la consolidación de la espectroscopía como herramienta fundamental en el análisis de compuestos bioactivos, como los del *Cordyceps militaris*. Después de **2018**, se percibe una **disminución progresiva** en las citas, seguida de una **caída abrupta posterior a 2020**, atribuible al **rezago natural en la citación de estudios recientes**, los cuales aún no han tenido el tiempo suficiente para acumular referencias (Figura 21).

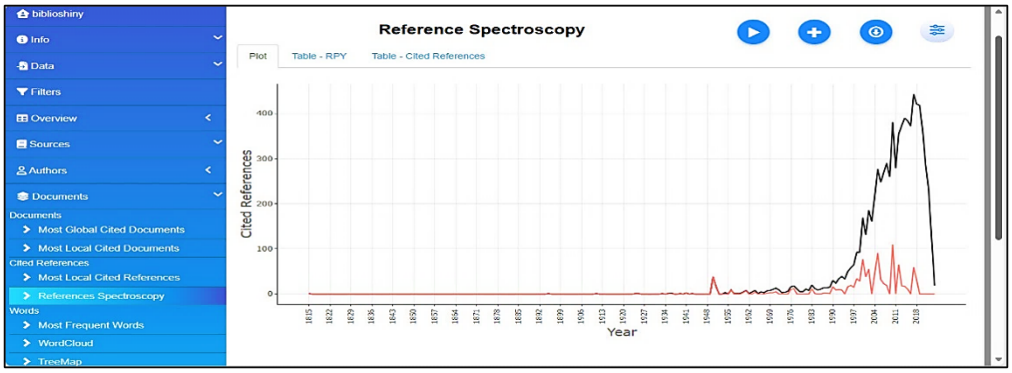


Figura 6.14: Figura 21

6.2.2.5 Modulo Clustering (Conceptual Structure)

Para el menú de *Conceptual Structure* concretamente en: ***Co-occurrence Network*** (traducido: Red de Coocurrencias), la red se encuentra claramente dividida en dos comunidades principales, identificadas por los colores rojo y azul. **La comunidad roja**, dominada por términos como cordycepin, Cordyceps militaris, metabolism y article, se orienta al estudio bioquímico y farmacológico del compuesto, mientras que **la comunidad azul** está asociada a modelos experimentales, destacando términos como animal experiment, human, mouse y cell line. Esta segmentación temática sugiere una dualidad en la línea de investigación: una centrada en la caracterización química y otra en los efectos biológicos en modelos preclínicos. El análisis de centralidad (como grado y betweenness) permitiría identificar términos puente como: nonhuman o controlled study, que conectan ambas comunidades (Figura 22).

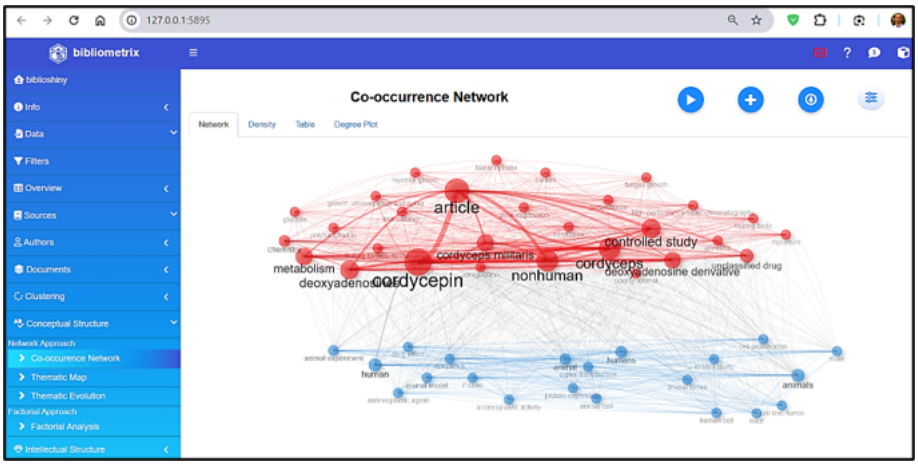


Figura 6.15: Figura 22

En el menú **Conceptual Structure**, específicamente en la opción **Factorial Analysis**, se presenta una representación bidimensional de los términos más relevantes identificados en el corpus bibliográfico analizado. En el **eje X (Dim 1)**, que explica el **50,18% de la variabilidad**, se agrupan términos estrechamente relacionados con **estudios experimentales in vivo e in vitro**, tales como *in vitro study*, *mouse*, *animal tissue* y *protein expression*, ubicados en el cuadrante inferior derecho. Este agrupamiento sugiere una fuerte carga temática asociada a **investigaciones biomédicas y farmacológicas**. Por su parte, el **eje Y (Dim 2)**, que explica un **9,68% adicional de la variabilidad**, concentra términos como *transcriptome* y *carbon*, vinculados a **estudios genéticos y metabólicos**, los cuales se encuentran espacialmente separados del resto de la nube léxica.

Esta segmentación espacial evidencia la existencia de **subdominios temáticos diferenciados** dentro del campo de estudio de los *Cordyceps* y sus derivados, revelando un **enfoque dual**: uno orientado hacia la **bioquímica y biotecnología del hongo**, y otro centrado en los **ensayos experimentales en organismos modelo** (Figura 23).

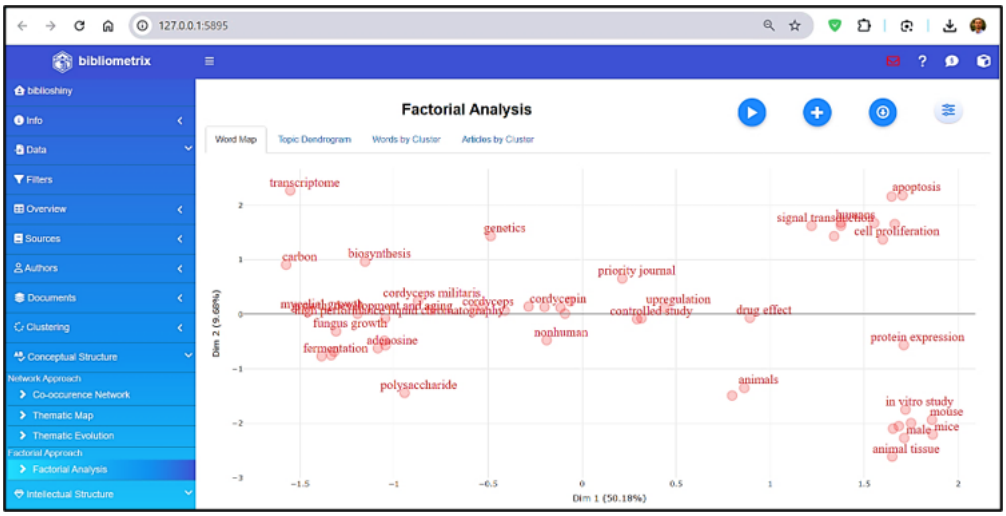


Figura 6.16: Figura 23

6.2.2.6 Social Structure

En **Social Structure**, específicamente en la opción **Collaboration Network**, se presenta un análisis basado en **redes de coautoría**, las cuales, según Aria y Cuccurullo (2017), permiten identificar **patrones de colaboración y productividad científica** entre los autores. En el contexto del estudio sobre *Cordyceps militaris*, la figura muestra una **red de colaboración entre investigadores**, donde **cada nodo representa un autor** y **las líneas o aristas** indican la existencia de coautoría en publicaciones científicas.

Se observan **múltiples grupos o comunidades de colaboración**, diferenciados por colores, lo que evidencia la existencia de **subgrupos de investigadores que trabajan conjuntamente de manera frecuente**. Algunos nodos, como **Li X** y **Vongsangnak W**, presentan un tamaño mayor, reflejando su **alto grado de colaboración** y centralidad dentro de la red. En contraste, otros autores aparecen más aislados, mostrando **menor interacción o participación** en proyectos conjuntos. Esta visualización permite comprender la **estructura social de la producción científica** sobre *Cordyceps militaris*, destacando los núcleos de liderazgo académico y las conexiones internacionales o institucionales más activas (Figura 24).

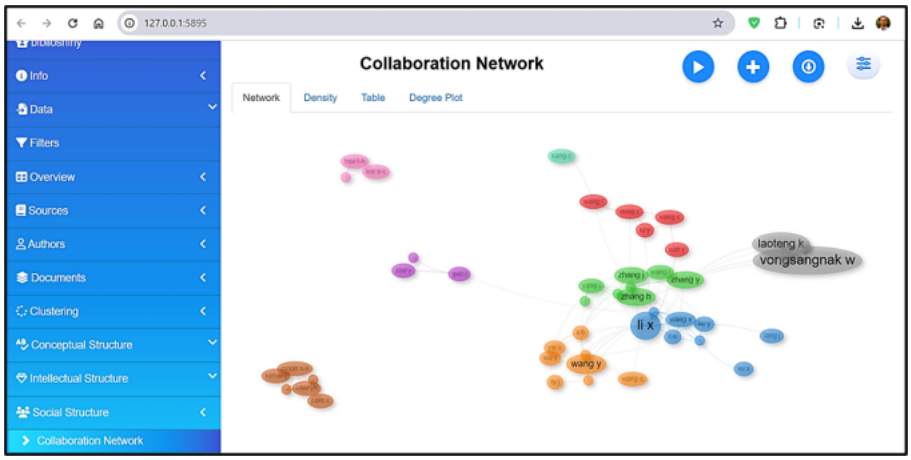


Figura 6.17: Figura 24

En la subcategoría **Social Structure → Countries' Collaboration World Map**, se visualizan las **redes internacionales de colaboración científica** sobre *Cordyceps* y su compuesto activo *cordycepina*.
El color azul indica el volumen de publicaciones —a mayor intensidad, mayor producción—, destacando **China** como el principal nodo y centro de cooperación. Las conexiones con **Estados Unidos, Alemania, Corea del Sur y Australia** evidencian un **patrón transcontinental de investigación**, que promueve la **transferencia de conocimiento y la consolidación de redes globales** en el estudio de hongos medicinales (Figura 25).

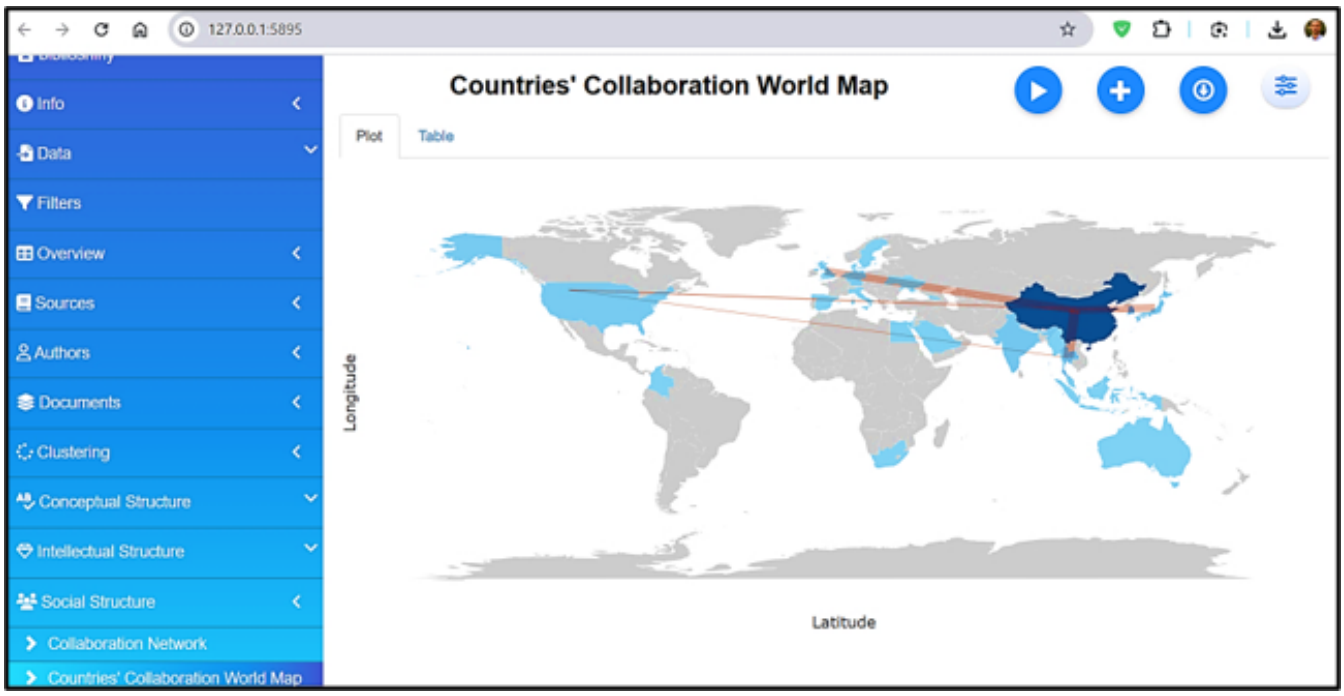


Figura 6.18: Figura 25

7 Capitulo 3 Generalidades del Diseño Experimental

El diseño experimental corresponde a una metodología científica y estadística destinada a planear, ejecutar y analizar pruebas controladas, con el propósito de obtener evidencia objetiva que responda a interrogantes sobre procesos o fenómenos específicos. El **diseño de experimentos (DOE)** se diferencia de la práctica empírica de prueba y error porque estructura el proceso investigativo bajo principios formales que permiten generar información confiable, optimizar recursos y reducir incertidumbre (Gutiérrez Pulido & Vara Salazar, 2012).

El DOE se da en ámbitos industriales y de investigación aplicada, los experimentos suelen realizarse para resolver problemas de calidad, mejorar procesos o comprobar hipótesis sobre materiales, condiciones de operación o métodos de trabajo. Sin embargo, cuando estas pruebas carecen de planeación rigurosa, se corre el riesgo de interpretar datos de manera subjetiva y desaprovechar el potencial de la variabilidad natural del sistema. Por ello, el DOE proporciona un marco que asegura resultados válidos y generalizables (Gutiérrez Pulido & Vara Salazar, 2012).

En cuanto a la terminología básica, conceptos como unidad experimental, tratamiento, factor controlable y no controlable, niveles de los factores, variable de respuesta, repetición y matriz de diseño, es requerido manejarlos. Estos términos constituyen la gramática operativa del diseño experimental, permitiendo estructurar adecuadamente las hipótesis y la recolección de datos. Además, se distingue entre error aleatorio y error experimental, resaltando la necesidad de minimizar y cuantificar ambos para garantizar validez estadística. Entre las etapas del diseño experimental, es incluyen:

- **Planeación:** formulación del problema, identificación de factores y niveles, selección de variables de respuesta y definición de objetivos.
- **Ejecución:** implementación del plan experimental bajo condiciones de control y aleatorización.
- **Análisis:** aplicación de métodos estadísticos, principalmente análisis de varianza (ANOVA), para estimar efectos principales e interacciones.
- **Interpretación:** extracción de conclusiones técnicas y toma de decisiones basadas en la evidencia.

Un aporte central son los principios básicos del DOE:

- **Aleatorización**, que asegura independencia de los errores y evita sesgos sistemáticos.
- **Replicación,** que incrementa la precisión de las estimaciones al cuantificar la variabilidad experimental.
- **Bloqueo**, que controla fuentes de variación no deseadas (turno, lote, operador), incrementando la potencia estadística del experimento.

Estos principios permiten estructurar experimentos que sean eficientes en costo y tiempo, pero robustos en cuanto a la validez de sus conclusiones.

La clasificación de diseños va desde los más simples (completamente al azar, bloques completos, cuadrados latinos) hasta los más complejos (factoriales, fraccionados, superficies de respuesta, diseños robustos). Se subraya que la selección depende de los objetivos, el número de factores, las restricciones prácticas y el tipo de información buscada. También se enfatiza que la decisión debe considerar tanto la significancia estadística como la significancia práctica, es decir, el impacto real de los resultados sobre el proceso o fenómeno bajo estudio (Gutiérrez Pulido & Vara Salazar, 2012).

7.0.1 2.1 Tipos de diseños experimentales

La selección de un diseño experimental depende de distintos factores que condicionan su pertinencia y aplicabilidad en cada situación. Entre los aspectos determinantes se encuentran: los objetivos que se persiguen con el estudio, la cantidad de factores que se desea analizar, el número de niveles que adoptará cada factor, los efectos que se pretende identificar en la relación causa-efecto y, finalmente, las restricciones de costo, tiempo y precisión que impone la investigación (Gutiérrez Pulido & Vara Salazar, 2012).

Estos elementos no actúan de forma aislada, ya que la modificación de cualquiera de ellos obliga generalmente a replantear el diseño a utilizar. En consecuencia, resultan fundamentales para guiar la clasificación de los diseños experimentales.

El **objetivo del experimento** constituye el criterio principal para diferenciar entre tipos de diseño, mientras que los demás factores funcionan como subcriterios de clasificación. Bajo esta perspectiva, los diseños pueden agruparse en varias categorías: aquellos orientados a la comparación de dos o más tratamientos; los que examinan el efecto de diversos factores sobre una o varias variables de respuesta; los que buscan establecer el punto óptimo de operación de un proceso; los que se enfocan en la optimización de mezclas; y finalmente, los dirigidos a lograr que un producto o proceso se mantenga estable frente a factores no controlables (Gutiérrez Pulido & Vara Salazar, 2012).

Así, la clasificación general de los diseños experimentales responde al objetivo central del estudio, y dentro de cada categoría se consideran elementos adicionales como el número de factores, los tipos de efectos a investigar y las restricciones prácticas que condicionan la ejecución.

7.0.2 Clasificación de los diseños experimentales

La siguente clasificación es tomada el libro de (Gutiérrez Pulido & Vara Salazar, 2012).

1. **Diseños para comparar dos o más tratamientos**
   Diseño completamente al azar
   Diseño de bloques completos al azar
   Diseño de cuadros latino y grecolatino

2. **Diseños para estudiar efectos de varios factores sobre una o más variables de respuesta**
   Diseños factoriales 2
   Diseños factoriales 3
   Diseños fraccionados 2
   Diseños anidados
   Diseños en parcelas divididas

3. **Diseños para la optimización de procesos**
   Modelo de primer orden
   Diseños factoriales 2 y 2
   Diseño de Plackett-Burman
   Diseño simplex
   Modelo de segundo orden
   Diseño de composición central
   Diseño de Box-Behnken
   Diseños factoriales 3 y 3

4. **Diseños robustos**
   Arreglos ortogonales (factoriales)
   Diseño con arreglos interno y externo

5. **Diseños de mezclas**
   Diseño simplex-reticular
   Diseño simplex con centroide
   Diseño sin restricciones
   Diseño axial

8 Capitulo 4 Diseño Completamente al Azar (DCA)

8.0.0.1 Problema

Introducción: La **antracnosis del banano** , causada por *Colletotrichum musae* (Berk. y M.A. Curtis) Arx, representa una problemática fitosanitaria de considerable relevancia económica en la industria bananera mundial, puesto que genera pérdidas postcosecha que oscilan entre el 10 y 80% debido al deterioro de la calidad visual del fruto, dicho patógeno desarrolla lesiones (formación de acérvulos) de coloración marrón oscuro a negro en el epicarpio del fruto, las cuales afectan la calidad visual del fruto (Vásquez-Castillo et al., 2019).

Tradicionalmente, el manejo de esta epifítia se ha fundamentado en la aplicación de fungicidas sintéticos como: tiabendazol, azoxystrobin y trifloxystrobin; no obstante, estas sustancias generan impactos ambientales adversos y residualidad ((Arias B., 2007), por ello, la búsqueda de alternativas de biocontrol sostenibles ha cobrado especial relevancia, particularmente mediante el uso de extractos fúngicos con propiedades antagónicas.

Metodología: El estudio se estructuró a partir de dos diseños experimentales: un **Diseño Completamente al Azar** para la evaluación de sustratos, y un **Diseño de Medidas Repetidas en el Tiempo** para la evaluación de la actividad inhibitoria.

8.0.0.2 Diseño 1: Sustratos de cultivo para *Penicillum* sp.

Se empleó un **Diseño Completamente al Azar** con los siguientes tratamientos: avena en hojuelas, maíz partido, semillas de cebada y arroz blanco. Se prepararon bolsas de polipropileno con cada sustrato, se inocularon con cinco discos de micelio de *Penicillum* sp. (0.5 mm de diámetro) y se incubaron de forma aleatorizada a 22 ± 2 °C durante ocho días. El experimento se realizó por quintuplicado, considerando cada bolsa como una repetición.

8.0.0.3 Diseño 2: Evaluación de la actividad inhibitoria

Se implementó un **Diseño de Medidas Repetidas en el Tiempo** para analizar el efecto de las concentraciones del extracto sobre dos variables de respuesta clave:

- **Porcentaje de Inhibición del Área de la Lesión (PIAL)**: Para evaluar la eficacia `in vivo`.
- **Porcentaje de Inhibición del Crecimiento Micelial (PICM)**: Para evaluar la eficacia `in vitro`.

Las variables independientes fueron las diferentes concentraciones del extracto y los testigos correspondientes, mientras que las variables de respuesta se midieron a lo largo del tiempo para observar la evolución de la inhibición.

Resultados: El maíz partido constituyó el sustrato óptimo para la producción conidial de *Penicillium digitatum*, alcanzando valores de Log_{10} 9,13 conidios/mL, seguido de la cebada Log_{10} 8,88 conidios/mL (**Figura 1**).

Figura 1.
Sustratos con Conidios de *Penicillium* sp.

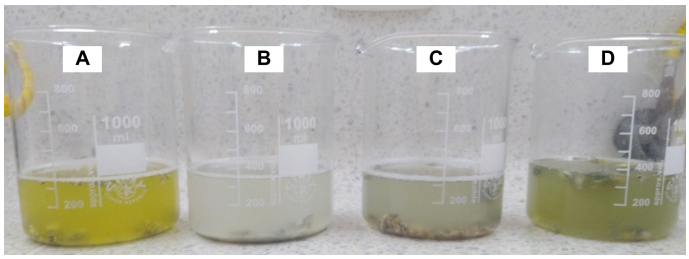


Figura 8.1: Nota: Dilución de conidios y sustrato, en solución tween80® 0,01%: Avena (A); Arroz (B); Cebada (C); Maíz Partido (D).

La evaluación in vitro reveló que las concentraciones de extracto crudo de 4,0 al 6,0% generaron Porcentajes de Inhibición del Crecimiento micelial (PICM) del 40 al 50 % respectivamente al quinto día después de la inoculación (ddi) **(Figura 2)**.

Figura 2.
Efecto de los tratamientos in vitro frente al crecmiento de *Colletotrichum musae.*

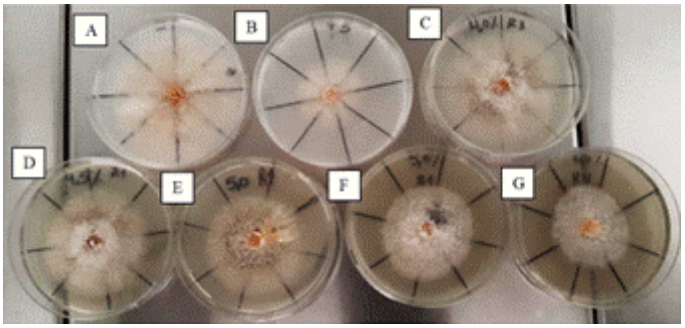


Figura 8.2: Nota: Prueba de inhibición in vitro de Colletotrichum musae, frente a diferentes tratamientos. (A) Testigo negativo; (B) Testigo positivo (Amistar a 60mg/100mL); (C) Extracto de *Penicillium* sp., al 4%; (D) Extracto de *Penicillium* sp., al 4,5%; (E) Extracto de *Penicillium* sp., al 5%; (F) Extracto de *Penicillium* sp., al 5,5%; (G) Extracto de *Penicillium* sp., al 6%.

Por otro lado, los ensayos in vivo evidenciaron una mayor eficacia del extracto crudo, donde las concentraciones de 8, 9, 10, 11, 12 y 13% generaron porcentajes de inhibición del área de la lesión (PIAL) de 60, 55, 70, 72, 77 y 80% respectivamente **(Figura 3)**, sugiriendo que Penicillium digitatum podría representar una alternativa viable para el manejo preventivo de la antracnosis del banano.

Figura 3.
Efecto in vivo de bananos infectados con *Colletotrichum musae* en los tratamientos.

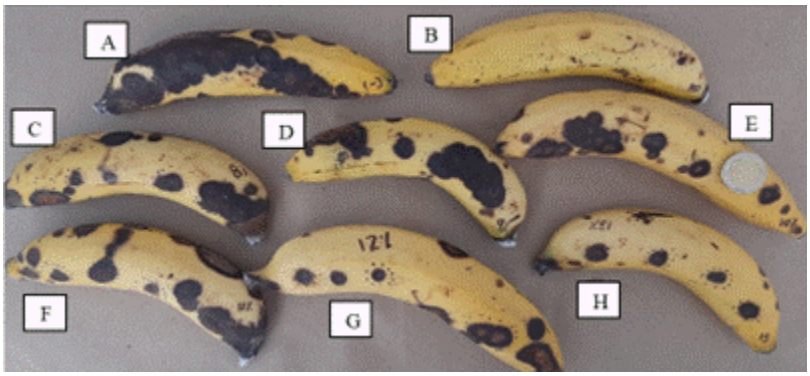


Figura 8.3: Nota: Experimento in vivo de los bananos infectados con 107 conidios de Colletotrichum musae, frente a tratamientos (A los 7 días de la inoculación). (A) Testigo negativo; (B) Azoxystrobin (Testigo positivo); Extractos de Penicillium sp. a (C) 8%; (D) 9%; (E) 10%; (F) 11%; (G) 12%; (H) 13%.

Para mayor información puede consultar: Mejía-Sarmiento, J. S. (2022). Evaluación de Extracto Crudo de Penicillium sp. para la Inhibición del Crecimiento in vitro e in vivo de Colletotrichum musae (Berk. y M. A. Curtis) Arx. Agente Causal de Antracnosis en Banano [Tesis de pregrado, Universidad de Santander UDES]. Repositorio Institucional UDES. https://repositorio.udes.edu.co/handle/001/8674

8.0.1 Estructura de la base de datos

La base de datos utilizada en este análisis corresponde a los resultados de un experimento agrícola que evalúa el comportamiento de cuatro cultivos diferentes bajo condiciones similares de manejo. La tabla contiene tres columnas principales:

Variable	Descripción
`Tratamiento`	Tipo de cultivo evaluado. Incluye cuatro niveles: Arroz, Avena, Cebada y Maíz.
`Repetición`	Número de repetición del tratamiento (del 1 al 4). Permite el análisis estadístico con replicación.
`Resultado`	Valor numérico correspondiente a la variable respuesta medida (por ejemplo, rendimiento en kg/ha).

Pasos para trabajar con R:

Especificar el directorio que me interesa donde se encuentra la base de datos.

> Antes e inciar
>
> R lee / (slash o division) y no el de Windows \
> En **R**, `setwd()` es una función que significa **"set working directory"** o "establecer el directorio de trabajo". Se utiliza para **definir la carpeta predeterminada** en la que R buscará archivos para leer y donde guardará archivos por defecto.
>
> Por ejemplo: setwd ("D:/OneDrive - Universidad de Santander/Material Docente 2025/CodigoR" ")

Lectura de datos

```
library(readxl)
```

```
Warning: package 'readxl' was built under R version 4.3.3
```

```r
DCA <- read_excel("C:/R-Proyectos/r-para-mi/data/dca.xlsx")

View(DCA)
attach(DCA)
names(DCA)
```

```
[1] "Tratamiento" "Repeticion"  "Resultado"
```

```r
str(DCA)
```

```
tibble [16 x 3] (S3: tbl_df/tbl/data.frame)
 $ Tratamiento: chr [1:16] "Arroz" "Arroz" "Arroz" "Arroz" ...
 $ Repeticion : num [1:16] 1 2 3 4 1 2 3 4 1 2 ...
 $ Resultado  : num [1:16] 8.76 8.74 8.72 8.72 8.39 ...
```

```r
summary(DCA$Resultado)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  8.341   8.635   8.792   8.775   8.954   9.141
```

Análisis de la Varianza - ANOVA

Cuando se desea saber si varios grupos (Ej. tratamientos) presentan diferencias reales en sus promedios, una de las herramientas estadísticas más utilizadas es el Análisis de la Varianza, conocido como ANOVA. Esta técnica permite examinar si los valores medios de tres o más grupos son lo suficientemente distintos como para concluir que no se trata de simples fluctuaciones aleatorias.

El enfoque de ANOVA se basa en comparar dos tipos de variación: por un lado, l**a variabilidad que se observa entre los distintos grupos**, y por otro, **la variabilidad que existe dentro de cada grupo individual**.

Si al analizar los datos se encuentra que la variación entre los grupos supera notablemente la que ocurre dentro de ellos, es razonable pensar que las diferencias en los promedios reflejan algo más que el azar. En cambio, si la variabilidad interna es más pronunciada, entonces es posible que las diferencias observadas no sean significativas y respondan a variaciones normales del comportamiento de los datos.

Código de R para ANOVA

```r
Anova<-aov(Resultado~Tratamiento, data=DCA)
summary(Anova)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Tratamiento  3 1.1794  0.3931   660.4 1.39e-13 ***
Residuals   12 0.0071  0.0006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretación: La prueba ANOVA muestra diferencias significativas entre los tratamientos ($p < 0.001$). El valor de F (660.4) indica que la variación entre tratamientos es mucho mayor que la variación dentro de los grupos, lo que sugiere que al menos uno de los tratamientos afecta significativamente el resultado.

Modelo Lineal

```r
modelo=lm(Resultado~(Tratamiento))
summary(modelo)
```

```
Call:
lm(formula = Resultado ~ (Tratamiento))

Residuals:
      Min        1Q    Median        3Q       Max
-0.038397 -0.016205  0.001983  0.012013  0.040116

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        8.73389    0.01220 715.921  < 2e-16 ***
TratamientoAvena  -0.35848    0.01725 -20.778 8.93e-11 ***
TratamientoCebada  0.12669    0.01725   7.343 8.94e-06 ***
TratamientoMaiz    0.39630    0.01725  22.970 2.75e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0244 on 12 degrees of freedom
Multiple R-squared:  0.994, Adjusted R-squared:  0.9925
F-statistic: 660.4 on 3 and 12 DF,  p-value: 1.393e-13
```

Interpretación: El modelo lineal confirma que el tratamiento influye significativamente en los resultados ($p < 0.001$). El tratamiento "Arroz" actúa como referencia, con una media estimada de 8.73. Comparado con este:

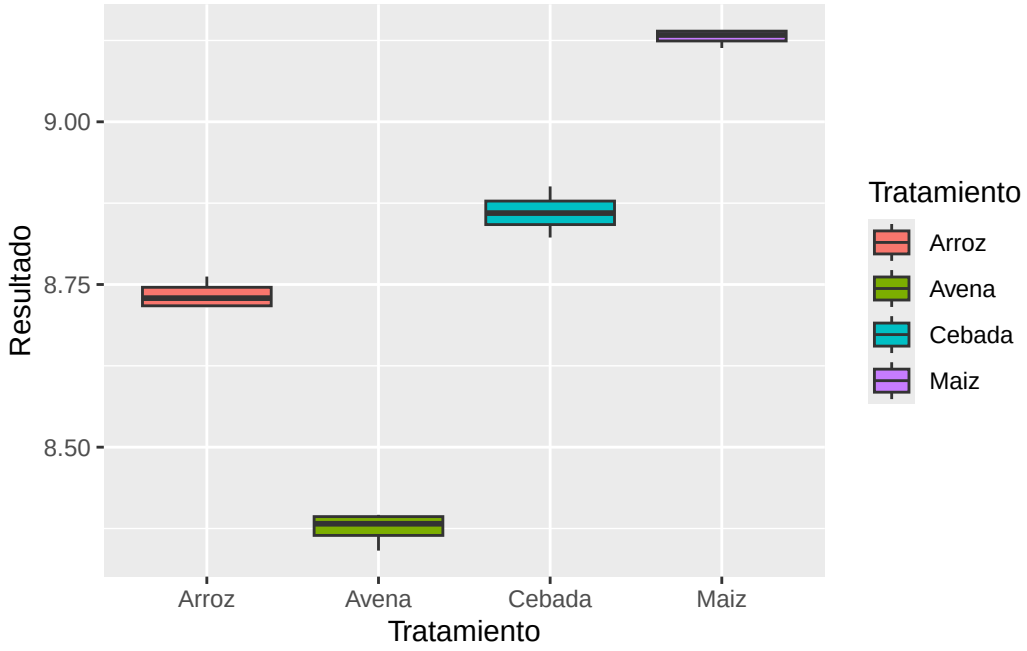
```
Avena presenta una media significativamente menor (-0.36, p < 0.001).

Cebada muestra un aumento moderado (+0.13, p < 0.001).

Maíz tiene el mayor incremento (+0.40, p < 0.001).
```

El modelo explica el 99.4% de la variabilidad en los datos ($R^2 = 0.994$), y el error estándar residual es bajo (0.0244), lo que indica un ajuste excelente.

Gráfico Boxplot

Se toma el Tratamiento para hacer un boxplot utilizando la variable "Resultado", pero primero se transformar en factor la variable Tratamiento:

```r
library(ggplot2)

DCA$Treatamiento<-factor(DCA$Tratamiento) #transformamos una variable númerica en un factor c
ggplot(DCA, aes(x = Tratamiento, y = Resultado, fill=Tratamiento)) +
  geom_boxplot()
```



Interpretación: Las diferencias en las medianas entre tratamientos son claras y consistentes con los resultados del ANOVA y del modelo lineal, lo que sugiere un efecto significativo del tipo de cultivo sobre la variable resultado.

Supuestos del diseño

Normalidad: Para verificar la normalidad de los residuos utilizaremos la prueba de Shapiro-Wilks cuyo script es el siguiente:

```
shapiro.test(residuals(Anova))
```

```
    Shapiro-Wilk normality test

data:  residuals(Anova)
W = 0.97944, p-value = 0.959
```

Interpretación: El test de Shapiro-Wilk aplicado a los residuos del modelo ANOVA devuelve un valor de p = 0.959, que es mucho mayor que 0.05. Esto indica que no hay evidencia estadística para rechazar la hipótesis nula de normalidad. Por lo tanto, se concluye que los residuos del modelo siguen una distribución normal, cumpliendo así uno de los supuestos fundamentales del análisis de varianza.

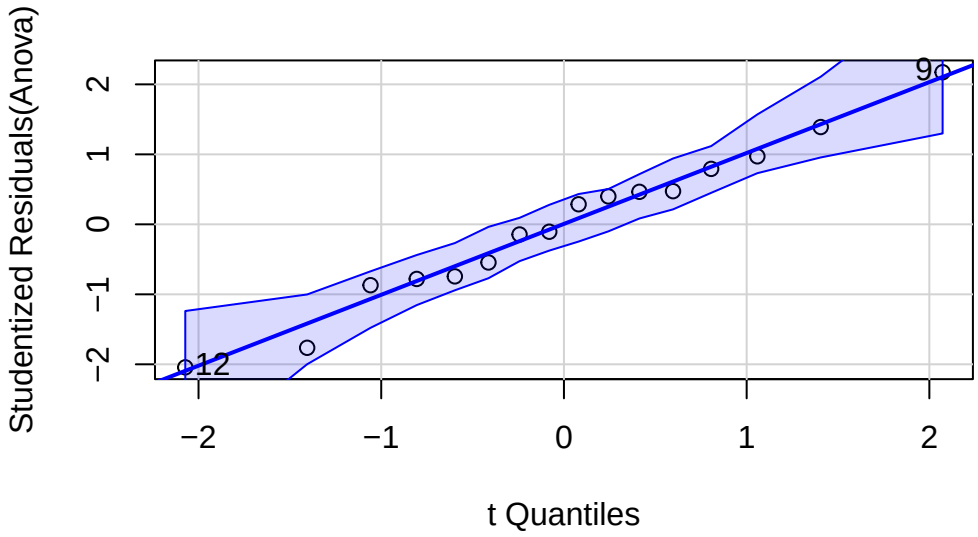
Gráficos para evaluar la normalidad

Para construir el gráfico QQ (QQ plot) y evaluar la normalidad de los datos, se utiliza la función correspondiente del paquete car. Si no está instalado previamente, es necesario instalar también el paquete auxiliar carData.

Instalación (si es necesario) install.packages ("car") install.packages ("carData") install.packages ("dplyr") install.packages ("purrr")

Cargar los paquetes (librerias)

```
library(car) #Grafico de QQ plot
library(carData)
library(dplyr)
library(purrr)

qqPlot(Anova)
```




```
[1]  9 12
```

Interpretación: El gráfico QQ muestra que los residuos estandarizados del modelo ANOVA se alinean adecuadamente con la línea diagonal, lo que indica que su distribución es aproximadamente normal. La mayoría de los puntos se ubican dentro de la banda de confianza, y no se observan desviaciones sistemáticas. Esta gráfica complementa el resultado del test de Shapiro-Wilk ($p = 0.959$), confirmando que se cumple el supuesto de normalidad de los residuos en el modelo.

Homocedasticidad: Para evaluar el supuesto de homogeneidad de varianzas entre los grupos (homocedasticidad), se aplicará la prueba de Bartlett, la cual es apropiada cuando los datos provienen de poblaciones aproximadamente normales. Esta prueba contrasta la hipótesis nula de igualdad de varianzas frente a la alternativa de varianzas diferentes. El procedimiento se implementa mediante el siguiente script:

```
bartlett.test(Resultado~Tratamiento, data=DCA)
```

```
    Bartlett test of homogeneity of variances

data:  Resultado by Tratamiento
Bartlett's K-squared = 2.2722, df = 3, p-value = 0.5179
```

Interpretación: Dado que el valor de p es mayor que 0.05 (p = 0.5179), no se rechaza la hipótesis nula. Por tanto, se asume que las varianzas entre los tratamientos son homogéneas, cumpliéndose este supuesto clave para el análisis de varianza y para la aplicación de pruebas a posteriori como LSD.

Pruebas aposteriori Para identificar diferencias específicas entre las medias de los tratamientos, una vez detectada significancia en el análisis de varianza, se aplicará una prueba de comparaciones múltiples a posteriori. En este caso, se empleará la técnica LSD (Least Significant Difference), que permite realizar comparaciones pareadas entre tratamientos asumiendo homogeneidad de varianzas.

La implementación de esta prueba requiere la carga del paquete agricolae, utilizando el siguiente script. Instalación si es necesario: install.packages("agricolae"). Carga del paquete: library(agricolae).

```
library(agricolae)
Grupos <- LSD.test(y = Anova, trt = "Tratamiento", group = T, console = T)
```

```
Study: Anova ~ "Tratamiento"

LSD t Test for Resultado

Mean Square Error:  0.0005953124

Tratamiento,  means and individual ( 95 %) CI

       Resultado        std r         se      LCL      UCL      Min      Max
Arroz   8.733890 0.02192214 4 0.01219951 8.707310 8.760471 8.715318 8.762183
Avena   8.375414 0.02519485 4 0.01219951 8.348834 8.401995 8.341039 8.395990
Cebada  8.860578 0.03330518 4 0.01219951 8.833998 8.887159 8.822181 8.900695
Maiz    9.130190 0.01251613 4 0.01219951 9.103609 9.156770 9.113429 9.140539
            Q25      Q50      Q75
Arroz  8.717232 8.729030 8.745688
Avena  8.364419 8.382314 8.393309
Cebada 8.842075 8.859719 8.878222
Maiz   9.124249 9.133395 9.139335

Alpha: 0.05 ; DF Error: 12
Critical Value of t: 2.178813

least Significant Difference: 0.03759044
```

```
Treatments with the same letter are not significantly different.

       Resultado groups
Maiz    9.130190      a
Cebada  8.860578      b
Arroz   8.733890      c
Avena   8.375414      d
```

Intrepretación: La prueba LSD reveló que los cuatro tratamientos presentan diferencias estadísticamente significativas entre sus medias. El tratamiento Maíz obtuvo el mayor rendimiento promedio, seguido por Cebada, Arroz y Avena, en ese orden descendente.

Otra opcion cuando cambiamos el argumento "group" a F(false), se interpreta a mi parecer de forma mas sencilla la diferencia entre las medias.A continuación, se presentan las pruebas de comparaciones múltiples a posteriori aplicadas al modelo de ANOVA ajustado. Se incluyen la prueba LSD, la prueba de Tukey y el test de Scheffé, las cuales permiten identificar diferencias estadísticamente significativas entre los tratamientos evaluados:

```r
Grupos<- LSD.test(y = Anova, trt = "Tratamiento", group = F, console = T)
```

```
Study: Anova ~ "Tratamiento"

LSD t Test for Resultado

Mean Square Error:  0.0005953124

Tratamiento,  means and individual ( 95 %) CI

       Resultado        std r         se      LCL      UCL      Min      Max
Arroz   8.733890 0.02192214 4 0.01219951 8.707310 8.760471 8.715318 8.762183
Avena   8.375414 0.02519485 4 0.01219951 8.348834 8.401995 8.341039 8.395990
Cebada  8.860578 0.03330518 4 0.01219951 8.833998 8.887159 8.822181 8.900695
Maiz    9.130190 0.01251613 4 0.01219951 9.103609 9.156770 9.113429 9.140539
           Q25      Q50      Q75
Arroz  8.717232 8.729030 8.745688
Avena  8.364419 8.382314 8.393309
Cebada 8.842075 8.859719 8.878222
Maiz   9.124249 9.133395 9.139335

Alpha: 0.05 ; DF Error: 12
Critical Value of t: 2.178813
```

```
Comparison between treatments means

               difference pvalue signif.        LCL         UCL
Arroz - Avena   0.3584760      0     ***  0.3208855  0.39606642
Arroz - Cebada -0.1266884      0     *** -0.1642788 -0.08909794
Arroz - Maiz   -0.3962994      0     *** -0.4338899 -0.35870901
Avena - Cebada -0.4851644      0     *** -0.5227548 -0.44757392
Avena - Maiz   -0.7547754      0     *** -0.7923659 -0.71718499
Cebada - Maiz  -0.2696111      0     *** -0.3072015 -0.23202064
```

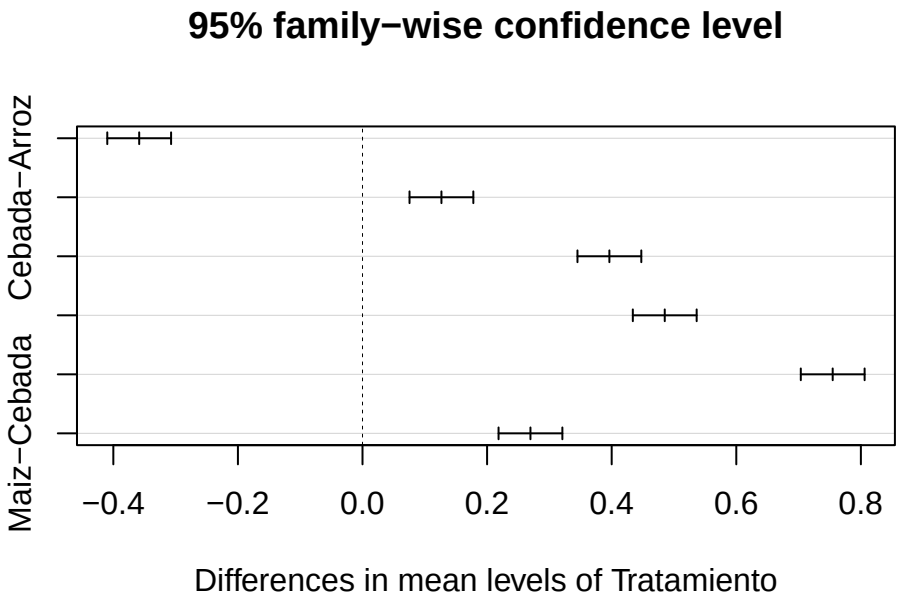
Interpretación: todas las diferencias entre tratamientos son altamente significativas ($p < 0.001$). Esto confirma que ninguno de los tratamientos comparte una media similar.

```
TukeyHSD(Anova)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Resultado ~ Tratamiento, data = DCA)

$Tratamiento
                    diff         lwr        upr   p adj
Avena-Arroz  -0.3584760 -0.40969759 -0.3072544 0.0e+00
Cebada-Arroz  0.1266884  0.07546677  0.1779100 4.6e-05
Maiz-Arroz    0.3962994  0.34507784  0.4475211 0.0e+00
Cebada-Avena  0.4851644  0.43394275  0.5363860 0.0e+00
Maiz-Avena    0.7547754  0.70355383  0.8059970 0.0e+00
Maiz-Cebada   0.2696111  0.21838947  0.3208327 0.0e+00
```

Interpretación: La prueba de Tukey también confirma diferencias estadísticamente significativas en todas las comparaciones, manteniendo control del error familiar. El gráfico generado muestra intervalos de confianza del 95% que no se solapan, lo que respalda visualmente los resultados.

```
plot(TukeyHSD(Anova))
```



Interpretación: El gráfico muestra los intervalos de confianza del 95 % para las diferencias de medias entre los tratamientos, ajustados por comparaciones múltiples (family-wise). Ninguno de los intervalos cruza la línea vertical en cero, lo cual indica que todas las comparaciones entre pares de tratamientos son estadísticamente significativas. La diferencia más grande se observa entre Maíz y Avena, mientras que la más pequeña, aunque significativa, es entre Cebada y Arroz. Este resultado es coherente con los análisis previos (ANOVA, LSD y Scheffé), y respalda que cada tratamiento tiene un efecto significativamente distinto sobre la variable "Resultado".

```
scheffe.test(Anova, "Tratamiento",console=TRUE)
```

```
Study: Anova ~ "Tratamiento"

Scheffe Test for Resultado

Mean Square Error  : 0.0005953124

Tratamiento,  means

       Resultado        std r         se      Min      Max      Q25      Q50
Arroz   8.733890 0.02192214 4 0.01219951 8.715318 8.762183 8.717232 8.729030
Avena   8.375414 0.02519485 4 0.01219951 8.341039 8.395990 8.364419 8.382314
```

```
Cebada  8.860578 0.03330518 4 0.01219951 8.822181 8.900695 8.842075 8.859719
Maiz    9.130190 0.01251613 4 0.01219951 9.113429 9.140539 9.124249 9.133395
            Q75
Arroz  8.745688
Avena  8.393309
Cebada 8.878222
Maiz   9.139335

Alpha: 0.05 ; DF Error: 12
Critical Value of F: 3.490295

Minimum Significant Difference: 0.05582762

Means with the same letter are not significantly different.

       Resultado groups
Maiz    9.130190      a
Cebada  8.860578      b
Arroz   8.733890      c
Avena   8.375414      d
```

Interpretación: A pesar de ser una prueba más conservadora, el test de Scheffé también encontró diferencias significativas entre todos los tratamientos. El análisis agrupó los tratamientos en distintos niveles.Mínima diferencia significativa (Scheffé): 0.0558. Valor crítico de F: 3.4903

Conclusión general Las tres pruebas aplicadas (LSD, Tukey y Scheffé) coinciden en que todos los tratamientos difieren significativamente entre sí. El tratamiento con mayor rendimiento fue Maíz, seguido por Cebada, Arroz y Avena, en orden descendente. Esto respalda la conclusión de que el tipo de tratamiento influye de manera significativa sobre la variable respuesta.

9 Capitulo 5 Diseño de Bloques Completamente al Azar (DBCA)

9.0.1 Problema

Introducción:

> Precaución
>
> Script: Análisis de un Diseño en Bloques Completamente al Azar (DBCA)
> Variables esperadas en la base:
> i. Bloque: factor que representa los bloques
> ii. Tratamiento: factor con los tratamientos a evaluar
> iii. Respuesta: variable cuantitativa a analizar #

9.1 Cargar Paquetes y librerías necesarias para el analisis de los datos.

9.2 Instalar Paquetes (solo una vez)

install.packages("readxl") install.packages("car") install.packages("agricolae") install.packages("lmtest") install.packages("ggplot2")

9.3 Cargar las siguientes librerias

```
library(readxl)
```

```
Warning: package 'readxl' was built under R version 4.3.3
```

```
library(car)
```

```
Loading required package: carData
```

```
Warning: package 'carData' was built under R version 4.3.3
```

```
library(agricolae)
```

```
Warning: package 'agricolae' was built under R version 4.3.3
```

```
library(lmtest)
```

```
Warning: package 'lmtest' was built under R version 4.3.2
```

```
Loading required package: zoo
```

```
Warning: package 'zoo' was built under R version 4.3.2
```

```

Attaching package: 'zoo'
```

```
The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

```
library(ggplot2)
```

9.4 Importar datos desde Excel

```
library(readxl)
DBCA <- read_excel("C:/R-Proyectos/r-para-mi/data/DBCA_Frijol/DBCA_frijol.xlsx")
```

9.5 Revisar estructura de los datos cargados

```r
View(DBCA)
attach(DBCA)
names(DBCA)
```

```
[1] "Tratamiento" "Bloque"      "Resultado"
```

```r
str(DBCA)
```

```
tibble [120 x 3] (S3: tbl_df/tbl/data.frame)
 $ Tratamiento: chr [1:120] "NPK Comercial" "NPK Comercial" "NPK Comercial" "NPK Comercial"
 $ Bloque     : chr [1:120] "A" "A" "A" "A" ...
 $ Resultado  : num [1:120] 51.5 49.6 51.9 54.6 49.3 ...
```

```r
summary(DBCA$Resultado)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  44.71   52.30   56.76   57.26   61.24   72.39
```

9.6 Convertir Variables a Factores

```r
DBCA$Tratamiento <- as.factor(DBCA$Tratamiento)
DBCA$Bloque <- as.factor(DBCA$Bloque)
```

9.7 ANOVA para Diseño de Bloques Completamente al Azar -DBCA-

```r
modelo <- aov(Resultado~Tratamiento + Bloque, data=DBCA)
summary(modelo)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
Tratamiento   3   3602  1200.6 153.749 <2e-16 ***
Bloque        2     11     5.4   0.688  0.505
Residuals   114    890     7.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

9.8 Verificación de supuestos

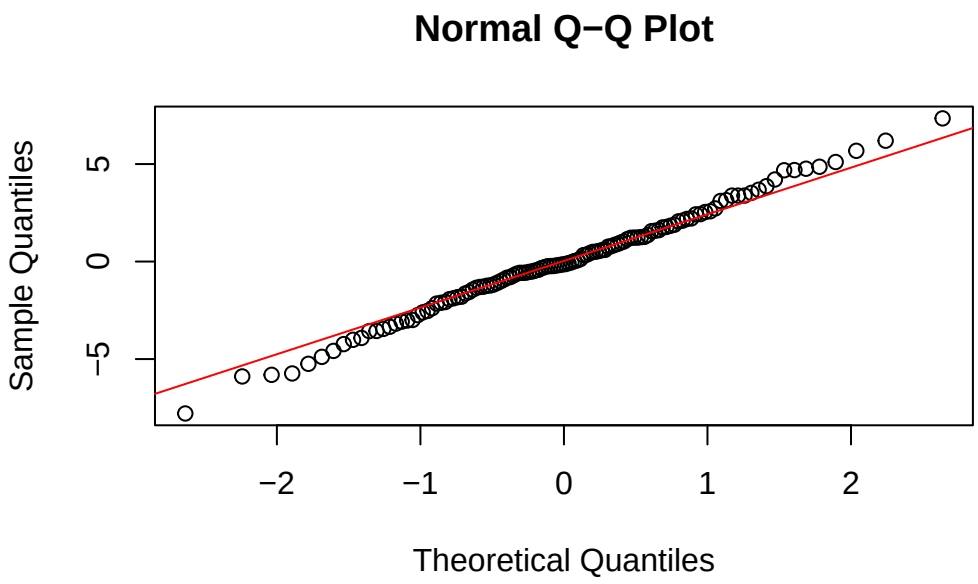
```
res <- residuals(modelo)
```

9.9 Normalidad de residuos (Shapiro-Wilk)

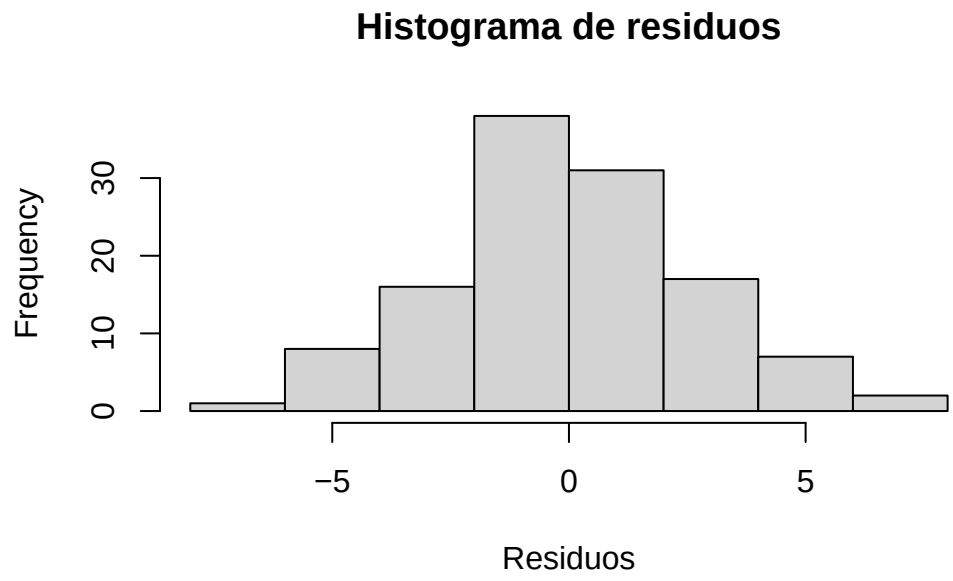
```
shapiro.test(res)
```

```
    Shapiro-Wilk normality test

data:  res
W = 0.99559, p-value = 0.9725
```

```
qqnorm(res)
qqline(res, col = "red")
```



```r
hist(res, main = "Histograma de residuos", xlab = "Residuos")
```



9.10 Homogeneidad de varianza

```r
bartlett.test(Resultado ~ Tratamiento, data = DBCA)
```

```
    Bartlett test of homogeneity of variances

data:  Resultado by Tratamiento
Bartlett's K-squared = 7.4355, df = 3, p-value = 0.05924
```

```r
car::leveneTest(Resultado ~ Tratamiento, data = DBCA)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   3  1.3691 0.2557
      116
```

```r
fligner.test(Resultado ~ Tratamiento, data = DBCA)
```

```
    Fligner-Killeen test of homogeneity of variances

data:  Resultado by Tratamiento
Fligner-Killeen:med chi-squared = 3.3125, df = 3, p-value = 0.3459
```

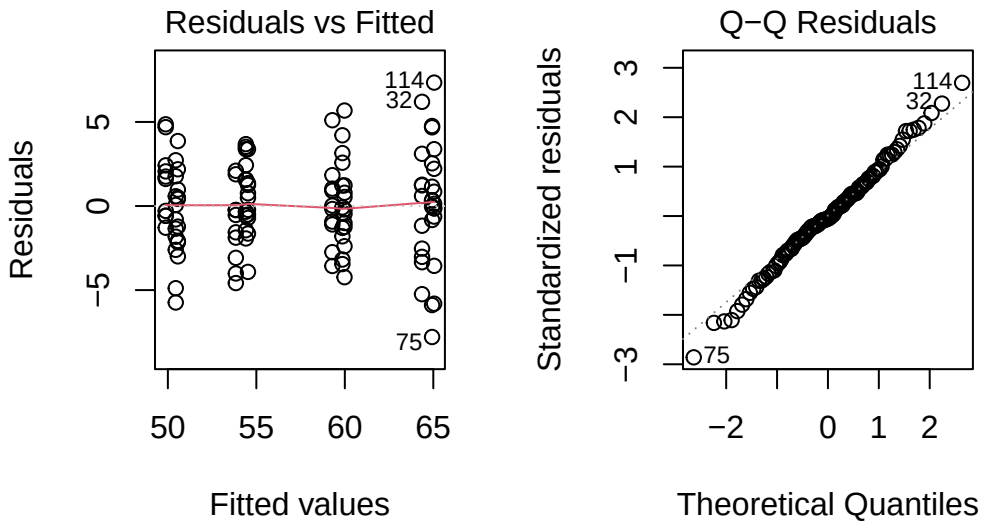
9.11 Independencia de residuos

```r
dwtest(modelo)
```

```
    Durbin-Watson test

data:  modelo
DW = 2.042, p-value = 0.4159
alternative hypothesis: true autocorrelation is greater than 0
```

9.12 Gráficos de diagnóstico del modelo

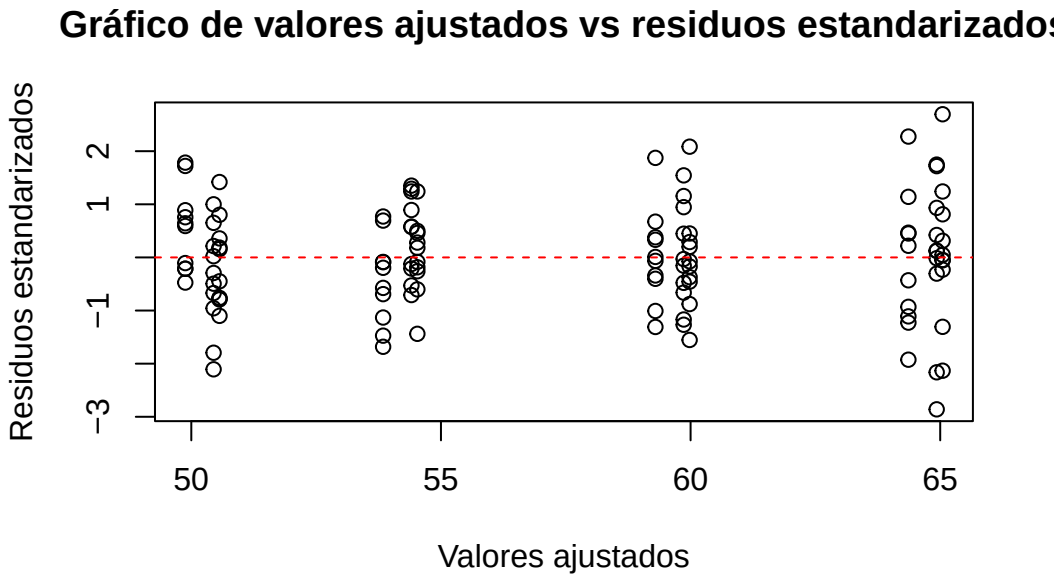
```r
# Gráficos de diagnóstico del modelo
par(mfrow = c(1, 2))  # Dividir la ventana en 2 gráficos (1 fila, 2 columnas)

plot(modelo, which = 1)  # Residuos vs valores ajustados (homogeneidad)
plot(modelo, which = 2)  # QQ-plot (normalidad)
```




```
par(mfrow = c(1, 1))  # Regresar a 1 gráfico por pantalla
```

9.13 Gráfico de valores ajustados vs residuos estandarizados

```
# Residuos estandarizados
resid_est <- rstandard(modelo)

# Valores ajustados predichos
val_ajus <- fitted(modelo)

plot(val_ajus, resid_est,
     main = "Gráfico de valores ajustados vs residuos estandarizados",
     xlab = "Valores ajustados",
     ylab = "Residuos estandarizados")
abline(h = 0, col = "red", lty = 2)
```



9.14 Comparaciones múltiples de medias

9.15 Tukey HSD (agrupamiento con letras)

```
out_Tukey <-HSD.test(modelo,"Tratamiento", group=TRUE, console=TRUE)
```

```
Study: modelo ~ "Tratamiento"

HSD Test for Resultado

Mean Square Error:  7.808829

Tratamiento,  means

                     Resultado      std  r        se   Min   Max     Q25
NPK + Bioinoculante   54.25933 2.395591 30 0.5101904 49.26 58.09 52.9100
NPK Comercial         50.29667 2.453096 30 0.5101904 44.71 54.74 48.5975
Orgánico + NPK        64.78000 3.621806 30 0.5101904 57.14 72.39 62.1700
```

```
Organico + NPK (50:50)  59.71300 2.488591 30 0.5101904 55.73 65.66 58.2425
                          Q50     Q75
NPK + Bioinoculante    54.045 55.9275
NPK Comercial          50.080 51.8625
Orgánico + NPK         65.070 66.9650
Organico + NPK (50:50) 59.480 61.0025

Alpha: 0.05 ; DF Error: 114
Critical Value of Studentized Range: 3.687325

Minimun Significant Difference: 1.881238

Treatments with the same letter are not significantly different.

                      Resultado groups
Orgánico + NPK          64.78000      a
Organico + NPK (50:50)  59.71300      b
NPK + Bioinoculante     54.25933      c
NPK Comercial           50.29667      d
```

9.16 Comparaciones múltiples con LSD

```r
out_LSD <- LSD.test(modelo, "Tratamiento", group = TRUE, console = TRUE)
```

```
Study: modelo ~ "Tratamiento"

LSD t Test for Resultado

Mean Square Error:  7.808829

Tratamiento,  means and individual ( 95 %) CI

                      Resultado      std  r        se      LCL      UCL   Min
NPK + Bioinoculante     54.25933 2.395591 30 0.5101904 53.24865 55.27002 49.26
NPK Comercial           50.29667 2.453096 30 0.5101904 49.28598 51.30735 44.71
Orgánico + NPK          64.78000 3.621806 30 0.5101904 63.76932 65.79068 57.14
Organico + NPK (50:50)  59.71300 2.488591 30 0.5101904 58.70232 60.72368 55.73
                        Max     Q25    Q50     Q75
NPK + Bioinoculante    58.09 52.9100 54.045 55.9275
```

```
NPK Comercial          54.74 48.5975 50.080 51.8625
Orgánico + NPK         72.39 62.1700 65.070 66.9650
Organico + NPK (50:50) 65.66 58.2425 59.480 61.0025

Alpha: 0.05 ; DF Error: 114
Critical Value of t: 1.980992

least Significant Difference: 1.429322

Treatments with the same letter are not significantly different.

                      Resultado groups
Orgánico + NPK          64.78000      a
Organico + NPK (50:50)  59.71300      b
NPK + Bioinoculante     54.25933      c
NPK Comercial           50.29667      d
```

9.17 Duncan

```r
out_Duncan <- duncan.test(modelo, "Tratamiento", group = TRUE, console = TRUE)
```

```
Study: modelo ~ "Tratamiento"

Duncan's new multiple range test
for Resultado

Mean Square Error:  7.808829

Tratamiento,  means

                      Resultado      std  r        se   Min   Max     Q25
NPK + Bioinoculante     54.25933 2.395591 30 0.5101904 49.26 58.09 52.9100
NPK Comercial           50.29667 2.453096 30 0.5101904 44.71 54.74 48.5975
Orgánico + NPK          64.78000 3.621806 30 0.5101904 57.14 72.39 62.1700
Organico + NPK (50:50)  59.71300 2.488591 30 0.5101904 55.73 65.66 58.2425
                          Q50     Q75
NPK + Bioinoculante    54.045 55.9275
NPK Comercial          50.080 51.8625
Orgánico + NPK         65.070 66.9650
```

```
Organico + NPK (50:50) 59.480 61.0025

Alpha: 0.05 ; DF Error: 114

Critical Range
       2        3        4
1.429322 1.504260 1.554074

Means with the same letter are not significantly different.

                      Resultado groups
Orgánico + NPK          64.78000      a
Organico + NPK (50:50)  59.71300      b
NPK + Bioinoculante     54.25933      c
NPK Comercial           50.29667      d
```

9.18 Bonferroni (usando pairwise.t.test)

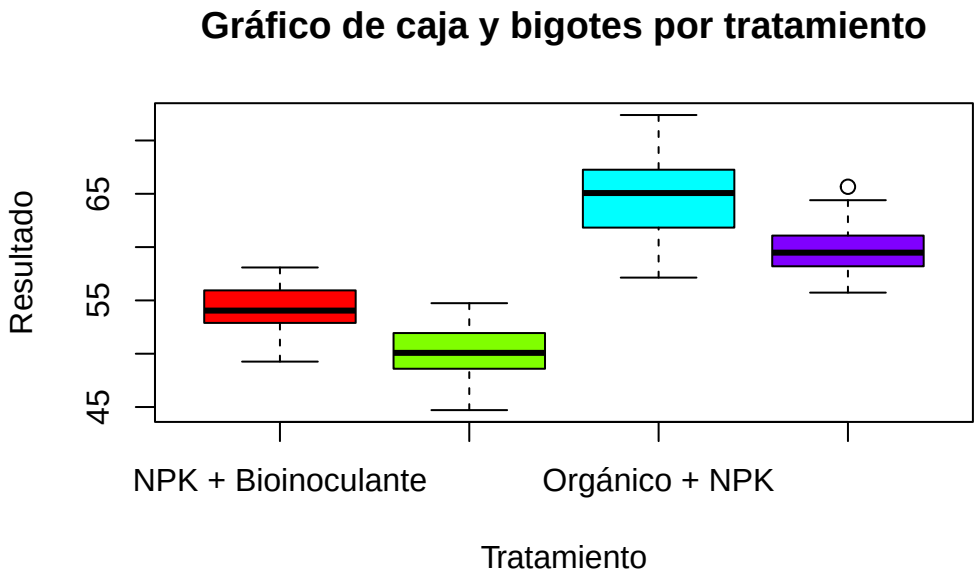
```r
out_Bonf <- pairwise.t.test(DBCA$Resultado, DBCA$Tratamiento, p.adjust.method = "bonferroni")
print(out_Bonf)
```

```
    Pairwise comparisons using t tests with pooled SD

data:  DBCA$Resultado and DBCA$Tratamiento

                      NPK + Bioinoculante NPK Comercial Orgánico + NPK
NPK Comercial         1.3e-06             -             -
Orgánico + NPK        < 2e-16             < 2e-16       -
Organico + NPK (50:50) 5.6e-11            < 2e-16       8.6e-10

P value adjustment method: bonferroni
```

9.19 Visualización de resultados

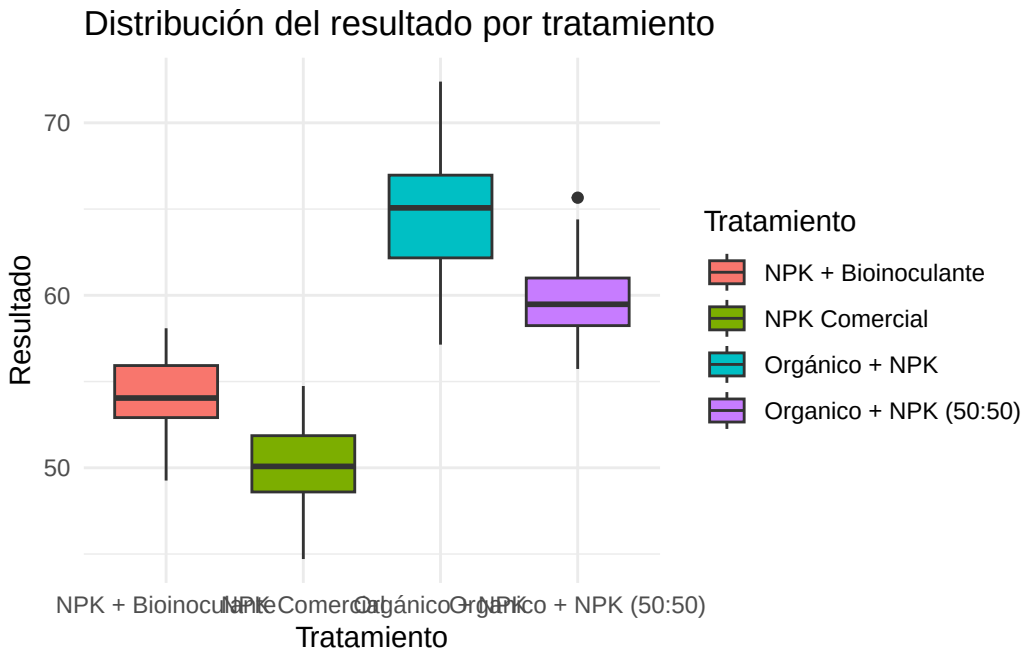
9.19.1 Caja y Bigote en R base

```r
boxplot(Resultado ~ Tratamiento, data = DBCA, main = "Gráfico de caja y bigotes por tratamien
```



9.20 Boxplot en ggplot2

```r
ggplot(DBCA, aes(x = Tratamiento, y = Resultado, fill = Tratamiento)) + geom_boxplot() + ther
```


Distribución del resultado por tratamiento
Resultado
70
60
50
NPK + Bioinoculante
NPK Comercial
Orgánico + NPK
Organico + NPK (50:50)
Tratamiento
Tratamiento
NPK + Bioinoculante
NPK Comercial
Orgánico + NPK
Organico + NPK (50:50)

10 Capitulo 6 Diseño longitudinal (ANOVA de medidas repetidas)

10.0.1 Problema

> **!** Importante
>
> **Metodología:** Se realizó un diseño de medidas repetidas en el tiempo, donde la variable independiente fue cada una de las concentraciones del extracto y los testigos; y la variable respuesta fueron: Porcentaje de Inhibición del Área de la Lesió (PIAL) y Porcentaje de Inhibición del Crecimiento Micelial (PICM).

10.1 Diseño: Tratamiento (factor entre sujetos) x Tiempo (factor intra-sujetos)

10.2 Función para instalar y cargar paquetes

```
use_pkg <- function(pkg){
  if (!require(pkg, character.only = TRUE)) {
    install.packages(pkg, dependencies = TRUE)
    library(pkg, character.only = TRUE)
  } else {
    library(pkg, character.only = TRUE)
  }
}
```

10.3 Instalar y cargar todos los paquetes necesarios

```
use_pkg("readxl")
```

```
Loading required package: readxl
```

```
Warning: package 'readxl' was built under R version 4.3.3
```

```
use_pkg("nlme")
```

```
Loading required package: nlme
```

```
Warning: package 'nlme' was built under R version 4.3.3
```

```
use_pkg("ggplot2")
```

```
Loading required package: ggplot2
```

```
use_pkg("emmeans")
```

```
Loading required package: emmeans
```

```
Warning: package 'emmeans' was built under R version 4.3.3
```

```
use_pkg("dplyr")
```

```
Loading required package: dplyr
```

```
Warning: package 'dplyr' was built under R version 4.3.3
```

```

Attaching package: 'dplyr'
```

```
The following object is masked from 'package:nlme':

    collapse
```

```
The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
use_pkg("multcomp")
```

```
Loading required package: multcomp

Warning: package 'multcomp' was built under R version 4.3.3

Loading required package: mvtnorm

Warning: package 'mvtnorm' was built under R version 4.3.3

Loading required package: survival

Warning: package 'survival' was built under R version 4.3.3

Loading required package: TH.data

Warning: package 'TH.data' was built under R version 4.3.3

Loading required package: MASS


Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select


Attaching package: 'TH.data'
```

```
The following object is masked from 'package:MASS':

    geyser
```

```
use_pkg("multcompView")
```

```
Loading required package: multcompView

Warning: package 'multcompView' was built under R version 4.3.3
```

10.4 IMPORTAR DATOS DESDE EXCEL

11 Importar datos

```
library(readxl)
DMRT <- read_excel("C:/R-Proyectos/r-para-mi/data/DMRT_Hongos/DMRT.xlsx")
```

```
View(DMRT)
attach(DMRT)
names(DMRT)
```

```
[1] "Tiempo"      "Tratamiento" "Repeticion"  "Resultado"
```

```
str(DMRT)
```

```
tibble [36 x 4] (S3: tbl_df/tbl/data.frame)
 $ Tiempo     : chr [1:36] "0 ddi" "0 ddi" "0 ddi" "10 ddi" ...
 $ Tratamiento: chr [1:36] "As 100 mg/L Pb" "As 100 mg/L Pb" "As 100 mg/L Pb" "As 100 mg/L Pb" "As 100 mg/L Pb" ...
 $ Repeticion : num [1:36] 1 2 3 1 2 3 1 2 3 1 ...
 $ Resultado  : num [1:36] 0 0 0 4.5 9.7 12.8 63.1 58.3 67.7 82.1 ...
```

```
summary(DMRT$Resultado)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.375  55.725  47.017  78.800  95.650
```

11.1 Convertir a factores

```
DMRT$Tiempo <- as.factor(DMRT$Tiempo)
DMRT$Tratamiento <- as.factor(DMRT$Tratamiento)
DMRT$Repeticion <- as.factor(DMRT$Repeticion)
```

11.2 MODELO MIXTO LINEAL (nlme)

```
modelo_mixto <- lme(Resultado ~ Tratamiento * Tiempo, random = ~1 | Repeticion, data = DMRT)
```

11.3 Resumen del modelo

```
anova(modelo_mixto)
```

```
                   numDF denDF  F-value p-value
(Intercept)            1    22 3324.775  <.0001
Tratamiento            2    22  100.341  <.0001
Tiempo                 3    22 1689.034  <.0001
Tratamiento:Tiempo     6    22   41.343  <.0001
```

11.4 COMPARACIONES POST-HOC (EMMEANS + SIDAK)

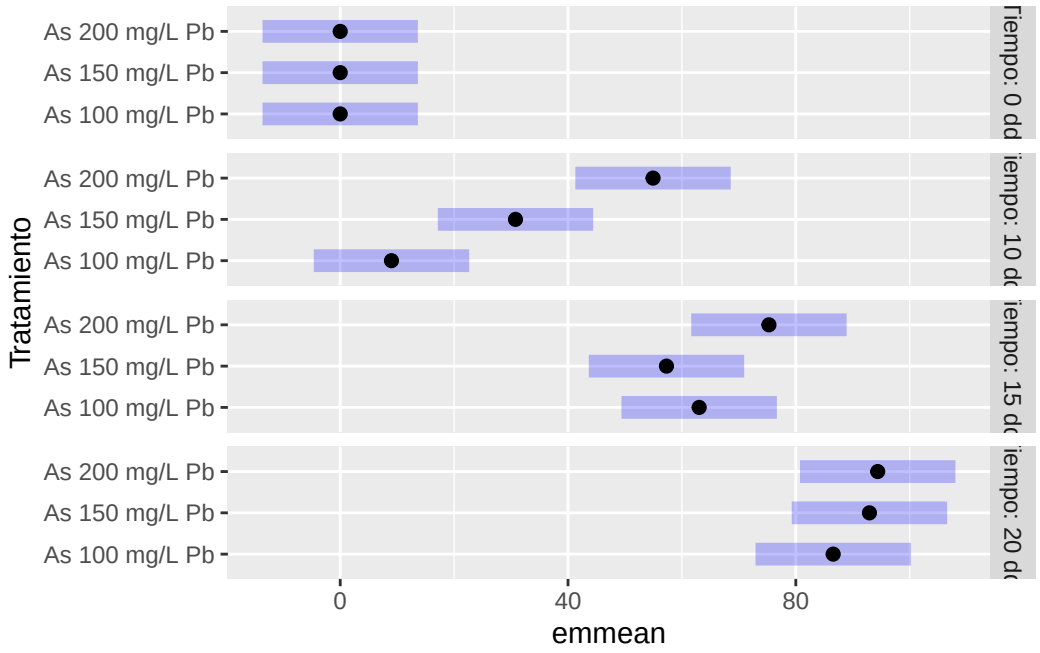
11.5 Tratamientos dentro de cada tiempo

```
# Comparaciones Post-Hoc (EMMEANS + SIDAK)
emmeans_trat_tiempo <- emmeans(modelo_mixto, ~ Tratamiento | Tiempo)

cld_trat_tiempo <- multcomp::cld(
  emmeans_trat_tiempo,
  adjust = "sidak",   # ajuste Sidak (Tukey no aplica en modelos mixtos)
  Letters = letters,
  alpha = 0.05
)

# Mostrar resultados y gráfico
print(cld_trat_tiempo)
```

```
Tiempo = 0 ddi:
 Tratamiento    emmean  SE df lower.CL upper.CL .group
 As 100 mg/L Pb    0.0 1.8  2   -13.64     13.6  a
```

```
 As 150 mg/L Pb    0.0 1.8  2   -13.64     13.6  a
 As 200 mg/L Pb    0.0 1.8  2   -13.64     13.6  a

Tiempo = 10 ddi:
 Tratamiento    emmean  SE df lower.CL upper.CL .group
 As 100 mg/L Pb    9.0 1.8  2    -4.64     22.6  a
 As 150 mg/L Pb   30.8 1.8  2    17.13     44.4   b
 As 200 mg/L Pb   54.9 1.8  2    41.29     68.6    c

Tiempo = 15 ddi:
 Tratamiento    emmean  SE df lower.CL upper.CL .group
 As 150 mg/L Pb   57.3 1.8  2    43.65     70.9  a
 As 100 mg/L Pb   63.0 1.8  2    49.39     76.7  a
 As 200 mg/L Pb   75.3 1.8  2    61.64     88.9   b

Tiempo = 20 ddi:
 Tratamiento    emmean  SE df lower.CL upper.CL .group
 As 100 mg/L Pb   86.6 1.8  2    72.92    100.2  a
 As 150 mg/L Pb   92.9 1.8  2    79.29    106.6   b
 As 200 mg/L Pb   94.4 1.8  2    80.74    108.0   b

Degrees-of-freedom method: containment
Confidence level used: 0.95
Conf-level adjustment: sidak method for 3 estimates
P value adjustment: sidak method for 3 tests
significance level used: alpha = 0.05
NOTE: If two or more means share the same grouping symbol,
      then we cannot show them to be different.
      But we also did not show them to be the same.
```

```
plot(cld_trat_tiempo)
```



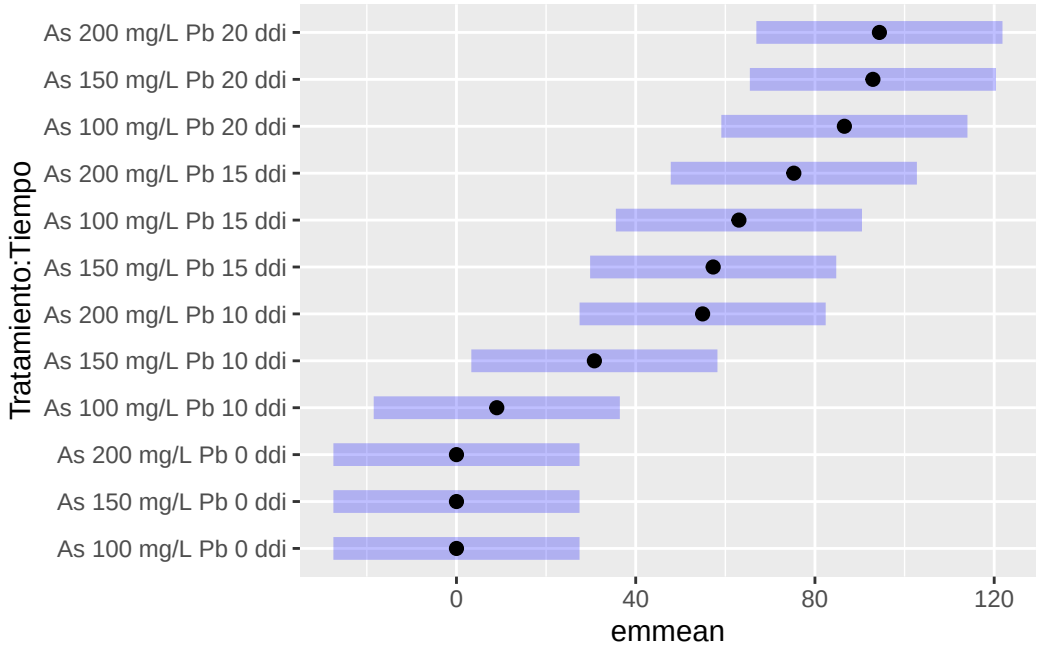
11.6 Interacción Tratamiento × Tiempo

```
emmeans_interaccion <- emmeans(modelo_mixto, ~ Tratamiento * Tiempo)

cld_interaccion <- multcomp::cld(emmeans_interaccion, adjust = "sidak", Letters = letters, al

print(cld_interaccion)
```

```
 Tratamiento    Tiempo emmean  SE df lower.CL upper.CL .group
 As 100 mg/L Pb 0 ddi     0.0 1.8  2   -27.46     27.5  a
 As 150 mg/L Pb 0 ddi     0.0 1.8  2   -27.46     27.5  a
 As 200 mg/L Pb 0 ddi     0.0 1.8  2   -27.46     27.5  a
 As 100 mg/L Pb 10 ddi    9.0 1.8  2   -18.46     36.5  a
 As 150 mg/L Pb 10 ddi   30.8 1.8  2     3.32     58.2   b
 As 200 mg/L Pb 10 ddi   54.9 1.8  2    27.47     82.4    c
 As 150 mg/L Pb 15 ddi   57.3 1.8  2    29.83     84.8    c
 As 100 mg/L Pb 15 ddi   63.0 1.8  2    35.57     90.5    c
 As 200 mg/L Pb 15 ddi   75.3 1.8  2    47.82    102.7     d
 As 100 mg/L Pb 20 ddi   86.6 1.8  2    59.10    114.0      e
 As 150 mg/L Pb 20 ddi   92.9 1.8  2    65.47    120.4      e
```

```
 As 200 mg/L Pb 20 ddi   94.4 1.8  2    66.92    121.8      e

Degrees-of-freedom method: containment
Confidence level used: 0.95
Conf-level adjustment: sidak method for 12 estimates
P value adjustment: sidak method for 66 tests
significance level used: alpha = 0.05
NOTE: If two or more means share the same grouping symbol,
      then we cannot show them to be different.
      But we also did not show them to be the same.
```

```
plot(cld_interaccion)
```



12 Capitulo 7 Uso de Inteligencia Artificial para la simulación de datos

12.1 Integración de la Inteligencia Artificial en la Simulación de Procesos Microbiológicos Industriales

La integración de herramientas de **inteligencia artificial (IA)** en la simulación de datos microbiológicos representa un cambio paradigmático en la investigación y el desarrollo de procesos biotecnológicos. Estas tecnologías permiten generar **datos sintéticos** que imitan comportamientos reales, facilitando la **validación de modelos estadísticos** antes de ejecutar experimentos en laboratorio (Amore & Philip, 2023). En este sentido, la IA impulsa la generación de datos experimentales, el desarrollo de modelos predictivos robustos y su aplicación a escala industrial.

Según (Wang et al., 2024), la IA ofrece soluciones innovadoras para el **análisis y simulación de datos**, al recrear dinámicas microbianas complejas y anticipar el efecto de cambios ambientales. Aunque no reemplaza la experimentación real, la complementa mediante escenarios virtuales que optimizan recursos y agilizan la toma de decisiones.

Su aplicación resulta especialmente valiosa en la **industria alimentaria**, donde contribuye al control del crecimiento bacteriano y a la mejora de la seguridad y calidad de los alimentos. Modelos computacionales basados en teorías de sistemas dinámicos han demostrado su eficacia al reproducir entornos industriales virtuales, reduciendo riesgos y optimizando parámetros de proceso (Melin & Castillo, 1996).

Asimismo, el **aprendizaje automático y profundo** ha revolucionado el procesamiento de datos microbiológicos, permitiendo la **caracterización de comunidades microbianas**, la interpretación de datos genómicos y el descubrimiento de metabolitos de interés biotecnológico [Kandilci et al. (2024)](Gurajala, 2024). Estas herramientas fortalecen la investigación en biotecnología industrial y microbiomas.

Otra innovación relevante es el desarrollo de **plataformas colaborativas** que democratizan el acceso a la simulación computacional, facilitando la construcción de modelos de crecimiento microbiano, producción metabólica y análisis de datos de secuenciación de nueva generación (Wasan, 2024). Esto promueve la cooperación entre el sector académico e industrial hacia bioprocesos más sostenibles.

De igual modo, (Fernández-Marín et al., 2025) destacan que los **modelos generativos** implementados en entornos como *Google Colab* permiten crear conjuntos de datos sintéticos que reproducen patrones reales, superando limitaciones derivadas de la variabilidad biológica y las restricciones de bioseguridad. La combinación de IA y herramientas estadísticas posibilita validar hipótesis, explorar escenarios hipotéticos y mejorar la **reproducibilidad experimental** en fermentaciones y control de contaminantes.

Finalmente, el futuro se orienta hacia la creación de **gemelos digitales** de bioprocesos: representaciones virtuales que integran datos biológicos, químicos y físicos para simular sistemas industriales completos. Esta línea emergente promete transformar la microbiología industrial al predecir el comportamiento de procesos con un nivel de detalle sin precedentes, impulsando la ingeniería de enzimas, la fermentación de precisión y la producción de biomoléculas de alto valor (Amore & Philip, 2023).

12.2 Diseño de *Prompts* para la Simulación de Datos en Microbiología Industrial

Desarrollar un *prompt* que permita generar un conjunto de datos simulados de crecimiento microbiano empleando una herramienta de Inteligencia Artificial (IA) generativa.

Objetivo: Aplicar principios de simulación de datos mediante IA para recrear escenarios experimentales propios de la microbiología industrial, obteniendo información sintética que pueda analizarse estadísticamente en **RStudio®**.

Instrucciones:

1. Copia y adapta el siguiente *prompt* en una herramienta de IA generativa como **ChatGPT**, **Claude**, **Copilot** o **Gemini**.
2. Sustituye los textos entre corchetes `[ ]` con tus propios valores o condiciones experimentales.
3. La simulación debe incorporar elementos **aleatorios** para reproducir la variabilidad natural de los experimentos biotecnológicos.

Primer modelo de Prompt

Genera un conjunto de datos simulados que imiten un experimento de crecimiento de [**nombre del microorganismo**] en condiciones industriales. Usa como base los siguientes parámetros experimentales: [**listar variables clave: temperatura, pH, concentración de sustrato, tiempo de incubación**]. Los datos deben incluir valores de densidad óptica (OD600), biomasa y metabolito producido.

El nuevo dataset debe tener al **menos 200 observaciones** y debe ser entregado en **formato CSV** para ser procesado en RStudio.
Una vez generado el dataset sintético, descárgalo en formato CSV y guárdalo en tu carpeta de trabajo de RStudio.

Segundo modelo de Prompt

Actúa como un investigador en Ingeniería de Bioprocesos y Microbiología Industrial. Genera un dataset simulado que represente la producción de amilasa por *Bacillus subtilis* en un biorreactor en lote.
Usa como variables de entrada:
(i)Temperatura (28–42 °C),
(ii) pH inicial (5.0–8.0),
(iii) Concentración de almidón soluble (5–50 g/L),
(iv) Aireación (0.5–2.0 vvm),
(v) Tiempo de fermentación (0–96 h).

Incluye como variables de salida:
(i) Biomasa (g/L),
(ii) Actividad de amilasa (U/mL),
(iii) Consumo de sustrato (%) y
(iv) pH final.
El dataset debe contener al menos 300 observaciones, estar en formato CSV, incluir variabilidad experimental (ruido aleatorio) y mantener relaciones biológicas plausibles (ejemplo: mayor sustrato y aireación → mayor biomasa y actividad enzimática hasta cierto límite).

Este dataset será procesado y analizado en RStudio mediante estadística descriptiva, regresión y visualización multivariante.

Tercer modelo de Prompt

Actúa como un microbiólogo industrial y bioestadístico. Genera un conjunto de datos simulados que representen curvas de crecimiento de *Escherichia coli* en medios líquidos bajo distintas condiciones.

Considera como variables de entrada:
(i) fuente de carbono (glucosa, lactosa, glicerol),
(ii) concentración inicial de sustrato (5–50 g/L),
(iii) temperatura (25–40 °C),
(iv) pH inicial (5.5–7.5) y

(v) tiempo de incubación (0–48 h).

Incluye como variables de salida:

(i) OD600,
(ii) fase de crecimiento y
(iii) velocidad específica de crecimiento (µ).

El dataset debe tener al menos 250 observaciones, exportarse en formato CSV, presentar curvas sigmoides biológicamente plausibles y reflejar variabilidad experimental realista.

Los datos se utilizarán en RStudio para modelar parámetros cinéticos y comparar el efecto de las condiciones de cultivo sobre el crecimiento bacteriano.

Cuarto modelo de Prompt

Actúa como un microbiólogo industrial y bioestadístico. Genera un conjunto de datos simulados que representen el Porcentaje de Inhibición del Área de la Lesión (PIAL) en frutos de banano tratados con extractos vegetales (Neem, Ajo, Jengibre y Canela) frente a *Colletotrichum musae*. Usa concentraciones de 1, 2.5, 5 y 10 %. Cada tratamiento debe contar con 5 repeticiones biológicas.
Emplea un diseño de Mediciones Repetidas en el Tiempo, con evaluaciones a los 3, 5, 7, 9 y 11 días después de la inoculación (ddi).
El archivo debe contener las siguientes columnas:
(i) Tratamiento,
(ii) Repetición,
(iii) días después de la inoculación,
(iv) PIAL.

Los valores de PIAL deben mostrar tendencias biológicas plausibles:
(i) El control (sin extracto) debe presentar progresión rápida de la lesión (PIAL cercano a 0%),
(ii) Los extractos vegetales deben mostrar distintos niveles de eficacia, algunos con inhibiciones superiores al 50%.

El dataset debe tener al menos 250 observaciones totales.

Exporta en formato CSV y asegúrate de que los datos reflejen tendencias biológicas plausibles: mayor concentración debe asociarse con mayor inhibición, cada extracto debe mostrar eficacia diferenciada y los valores deben incluir variabilidad experimental.

Referencias

Amore, A., & Philip, S. (2023). Artificial intelligence in food biotechnology: trends and perspectives. *Frontiers in Industrial Microbiology*. https://doi.org/10.3389/finmi.2023.1255505

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Arias B., C. L. (2007). Control Químico de la Antracnosis del Mango (Mangifera indica L.) en pre y postcosecha. *Bioagro*, *19*(1), 19-25. http://www.ucla.edu.ve/bioagro/Rev19(1)/3.%20Control%20qu%C3%ADmico%20de%20la%20antracnosis.pdf

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2021). *shiny: Web Application Framework for R*. https://CRAN.R-project.org/package=shiny

Fernández-Marín, M. Á., Montero-Murillo, J. R., & González-Tolmo, D. (2025). Caso de estudio sobre simulación de datos para investigaciones académicas mediante Inteligencia Artificial Generativa y Google Colab. *Revista Mexicana de Investigación e Intervención Educativa*, *4*(S1), 18-26. https://doi.org/10.62697/rmiie.v4iS1.143

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3.ª ed.). Sage Publications. https://uk.sagepub.com/en-gb/eur/an-r-companion-to-applied-regression/book246125

Gurajala, S. (2024). Artificial intelligence (AI) and medical microbiology: A narrative review. *Indian Journal of Microbiology Research*. https://doi.org/10.18231/j.ijmr.2024.029

Gutiérrez Pulido, H., & Vara Salazar, R. de la. (2012). *Análisis y diseño de experimentos* (3.ª ed.). McGraw-Hill/Interamericana Editores, S.A. de C.V.

Kandilci, M., Yakıcı, G., & Kayar, M. (2024). Artificial intelligence and microbiology. *Experimental and Applied Medical Science*. https://doi.org/10.46871/eams.1458704

Lahti, L., & Shetty, S. (2017). *microbiome R package*. http://microbiome.github.io/microbiome

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Melin, P., & Castillo, O. (1996). Modelling and simulation for bacteria growth control in the food industry using artificial intelligence. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 676-681. https://www.witpress.com/elibrary/wit-transactions-on-information-and-communication-technologies/16/13768

Mendiburu, F. (2020). *agricolae: Statistical Procedures for Agricultural Research*. https://CRAN.R-project.org/package=agricolae

Mohammadi, R., Ghomi, S. M. T. F., & Nazari, F. (2019). The application of R software for the assessment of microbial fermentation processes. *Journal of Microbiological Methods*, *156*, 54-58. https://doi.org/10.1016/j.mimet.2018.12.003

Navarro, D. J. (2015). *Learning Statistics with R: A tutorial for psychology students and other beginners* (Versión 0.5). University of Adelaide. https://learningstatisticswithr.com/

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *vegan: Community ecology package*. https://CRAN.R-project.org/package=vegan

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2025). *nlme: Linear and nonlinear mixed effects models*. https://CRAN.R-project.org/package=nlme

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ritz, C., & Streibig, J. C. (2005). Bioassay analysis using R. *Journal of Statistical Software*, *12*(5), 1-22. https://doi.org/10.18637/jss.v012.i05

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Vásquez-Castillo, W., Racines-Oliva, M., Moncayo, P., Viera, W., & Seraquive, M. (2019). Calidad del fruto y pérdidas postcosecha de banano orgánico (Musa acuminata) en el Ecuador. *Enfoque UTE*, *10*(4), 57-66. https://doi.org/10.29019/enfoque.v10n4.545

Wang, X. W., Wang, T., & Liu, Y. Y. (2024). Artificial intelligence for microbiology and microbiome research. *arXiv*. https://doi.org/10.48550/arXiv.2411.01098

Wasan, R. K. (2024). The role of artificial intelligence (AI) in microbiology laboratories for diagnosis of microorganisms: A review study. *International Journal of Life Sciences Biotechnology and Pharma Research*. https://doi.org/10.69605/ijlbpr_13.8.2024.66

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2.ª ed.). Springer. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2015). *readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl

Wickham, H., & Grolemund, G. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media. https://r4ds.had.co.nz

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for

visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28-36. https://doi.org/10.1111/2041-210X.12628

Zhou, B., Xiao, J. F., Tuli, L., & Ressom, H. W. (2012). LC-MS-based metabolomics. *Molecular BioSystems*, *8*(2), 470-481. https://doi.org/10.1039/c1mb05350g