

ESTADÍSTICA

Aplicada a las Ciencias Ambientales y
de la Vida:

Una perspectiva Computacional con
Excel®

Dr. José Ángel Gaspar - Génico



Herramienta clave para Biólogos • Ingenieros Ambientales
• Gestores



Primera edición, 2026		
D. R. © 2026, José Ángel Gaspar Genico.		
	Villahermosa, Centro, Tabasco, México. Domicilio del titular de los derechos patrimoniales para efectos legales y fiscales, conforme a la normativa del INDAUTOR.	
Todos los derechos reservados. Esta obra se publica en formato impreso y digital. Queda prohibida su reproducción total o parcial por cualquier medio, electrónico, mecánico, óptico, de grabación o cualquier otro, sin el permiso previo y por escrito del titular, conforme a la Ley Federal del Derecho de Autor. La obra se encuentra registrada ante el Instituto Nacional del Derecho de Autor (INDAUTOR).		
Para su publicación esta obra ha sido dictaminada por el sistema académico de “pares ciegos”. Los juicios expresados son responsabilidad del autor.		
Diseño portada: Stephanie Pérez Alamilla Fotografía de autor: José Ángel Gaspar Genico		
ISBN: 978-970-96059-0-7 [versión impresa] ISBN: 978-970-960xx-x-x [versión digital]		
Hecho en Villahermosa, Centro, Tabasco, México.		
Aviso sobre marcas registradas		
Microsoft y Excel son marcas registradas de Microsoft Corporation. Su mención en esta obra tiene fines exclusivamente descriptivos y educativos. Este libro no está afiliado ni avalado por Microsoft Corporation. Todas las marcas registradas pertenecen a sus respectivos propietarios.		

Dedicatoria

Al creador del Universo, por su maravillosa creación de los números

A mi padre Eduardo y a mi madre Hilda

A toda mi familia

A mis amigos

Estadística aplicada a las ciencias ambientales y de la vida: una perspectiva computacional con Excel.

Agradezco a

A mis maestros y alumnos de Estadística

A mis tesisas

A mis compañeros de trabajo

A la Universidad Juárez Autónoma de Tabasco

A los revisores de este libro, grandes expertos y especialistas en Estadística



Prologo

El libro "Estadística aplicada a las ciencias ambientales y de la vida: una perspectiva computacional con Excel" llena un espacio necesario en la práctica docente que será de utilidad al estudiante, al académico, así como a los profesionales de ciencias ambientales y de la vida que requieran una guía aplicada, completa, rápida y contextualizada que minimice los errores en los trabajos futuros. Con el autor de la obra coincidimos en que la estadística no debe percibirse como un obstáculo matemático sino como un lenguaje para comprender la naturaleza y tomar decisiones informadas. Cada capítulo busca acompañarte desde el registro correcto de datos hasta el análisis y la interpretación de resultados mediante ejercicios con datos comunes en los trabajos de investigación ya que nace de la experiencia docente y de investigación compartida en las ciencias ambientales y de la vida. El objetivo central es que desarrolles criterio científico, pensamiento crítico y autonomía analítica. Si logras conectar los números con procesos ecológicos, ambientales y sociales, este libro habrá cumplido plenamente su propósito formativo en tu trayectoria Universitaria y profesional. Estoy seguro que este libro seguramente será una herramienta indispensable en tu trabajo diario que te permitirá alcanzar tus metas.

Dr. Luis Manuel Vargas Villamil
Investigador en Modelación Matemática

Índice

Consideraciones generales..... 1

1. Matemáticas y la Estadística.....2

 1.1. Definición de la estadística.....3

 1.2. Estadística descriptiva4

 1.3. Estadística inferencial5

 1.4. Teoría del muestreo6

 1.4.1. Qué es el Muestreo7

 1.4.2. El Universo y la Población.....7

 1.4.3. La Muestra7

 1.4.4. Aleatorio7

2. Que son las variables y los datos.....8

 2.1. Elección de variable8

 2.2. Recopilación de información.....8

 2.3. Tipos de datos y su relevancia.....9

 2.4. Clasificación de datos: Escalas de Medición11

 2.5. Ordenamiento de datos13

3. Uso de Excel® 2016 para el Análisis de Datos Estadísticos14

 3.1. Que es Excel® y que es una base de datos14

 3.2. Interfaz de Excel® 2016: celdas, filas, columnas, hojas14

 3.3. Captura y formato de datos en Excel: números, texto, fechas.....16

 3.4. Registro de datos numéricos17

 3.5. Ordenar datos.....18

 3.6. Filtrar datos en Excel®19

3.7. Operaciones básicas en Excel®: suma, resta20

3.8. Uso de funciones estadísticas en Excel®21

3.9. Uso de Funciones.....22

4. Estadística descriptiva23

4.1. Estadística descriptiva numérica24

4.2. Cómo elegir el tipo de Gráfico27

4.2.1. Gráfico de Columnas29

4.2.2. Gráfico de Líneas.....30

4.2.3. Gráfico Circular31

4.2.4. Gráfico de Barras.....32

4.2.5. Gráfica de Áreas33

4.2.6. Gráfica X Y (dispersión)34

4.2.7. Gráfica de Mapa35

4.2.8. Gráfica de Cotizaciones.....36

4.2.9. Gráfica de Superficie37

4.2.10. Gráfica Radial38

4.2.11. Gráfica de rectángulos.....39

4.2.12. Gráfica de Proyección solar.....40

4.2.13. Gráfica de Histograma41

4.2.14. Gráfica de Cajas y bigotes42

4.2.15. Gráfica de Cascada43

4.2.16. Gráfica de Embudo44

4.2.17. Gráfica de Combinado, ejemplo Climograma45

4.2.17. Gráfica de Combinado, ejemplo Diagrama de Balance Hídrico, para detectar periodos de sequia.....46

5. Estadística inferencial47

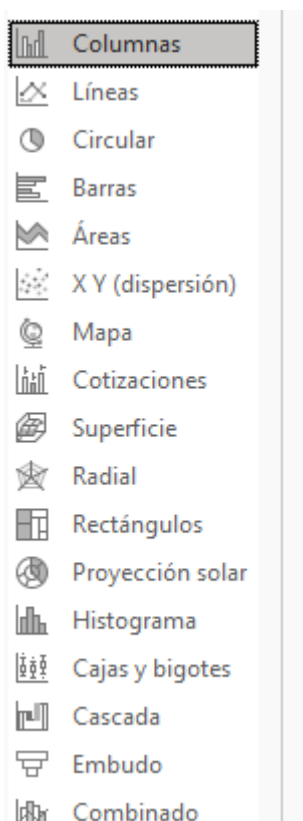
Estimación de parámetros	47
Intervalos de confianza	47
Incertidumbre	47
Pruebas de hipótesis y nivel de significancia	48
La Hipótesis nula (H_0) y alternativa (H_a)	48
El valor p , su relación con α y error tipo I	50
5.1. Estadística paramétrica	54
5.1.1. Normalidad	54
5.1.2. Homocedasticidad	54
5.1.3. Selección de Técnicas Estadísticas para analizar datos	55
5.1.4. Técnicas estadísticas para Describir	56
5.1.5. Técnicas Estadísticas Paramétricas para comparar grupos	63
5.1.6. Técnicas Estadísticas para relacionar variables.....	85
5.1.7. Técnicas estadísticas para Predecir	92
5.1.8. Técnicas estadísticas para Descubrir patrones especiales	101
5.2. Estadística No Paramétrica	112
Referencias.....	113
Índice Analítico	119

Consideraciones generales

Usa este libro para realizar tus anotaciones de estadística, dentro de las páginas.

"Microsoft Excel® es una herramienta desarrollada por Microsoft Corporation. El uso de este nombre en este libro es con fines educativos, sin fines de lucro ni relación comercial con la empresa."

1



PRESENTACION

1. Matemáticas y la Estadística

Las matemáticas son el lenguaje con que podemos comprender cómo el Universo se comunica e interactúa entre todos sus componentes. El Biólogo, el Ingeniero Ambiental o el Gestor Ambiental que conozcan su lenguaje podrán apreciar sus secretos, su funcionamiento, comunicarlos y usarlos en beneficio de la sociedad y el medio ambiente.

Las matemáticas estudian números, figuras y patrones usando un lenguaje que todos pueden entender. Su objetivo es encontrar reglas y patrones que nos ayuden a resolver problemas en la vida diaria, en la escuela y en el trabajo.

Las matemáticas se apoyan de diferentes disciplinas las cuales interrelacionadas nos ayudan a comprender desde una forma sencilla hasta formas muy complejas. Dentro de estas disciplinas se encuentran: Estadística, Probabilidad, Álgebra, Geometría, Trigonometría, Aritmética, Topología, Análisis, Combinatoria, Matemáticas recreativas como las principales.

1.1. Definición de la estadística

La estadística se aplica en todas las disciplinas científicas cuando necesitamos incrementar el conocimiento o resolver problemas mediante la recopilación y análisis de datos para comprender fenómenos y tomar decisiones (Mendenhall *et al.* 2010; Wackerly, *et al.* 2010).

3



“La estadística es la ciencia que recopila, organiza, analiza e interpreta datos para la toma de decisiones” (Freedman, Pisani & Purves, 2007).



Rta	Nominal	Ordinal	Intervalo (°C)	Razón (g)
1	Tomate rojo	Madurez	Si	100
2	Cherry amarillo	Dulzura	Si	20
3	Cherry naranja	Presencia	Si	20
4	Cherry rojo	Calidad	Si	20
5	Cherry negro	Rarez	Si	20
6	Guacamole	Tamaño	Si	120
7	Tomate verde	Acidez	Si	100
8	Tomate rojo	Firmeza	Si	200
9	Juza rojo	Presentación	Si	10
10	Juza amarillo	Conservación	Si	10

Recopila

organiza

Analiza

Interpreta

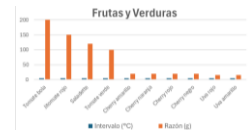


Figura 1.1. Etapas fundamentales del análisis estadístico.

La estadística ayuda en Biología a estudiar poblaciones de especies y funcionamiento de ecosistemas, mientras que en Ingeniería Ambiental permite evaluar contaminantes en agua, aire y suelo, que permiten identificar la salud del ambiente.

En ambas disciplinas, los datos estadísticos revelan patrones y relaciones que explican cómo funcionan los organismos y el medio ambiente. Y esto permite dar fundamento a la Gestión para la rehabilitación.

1.2. Estadística descriptiva

La estadística descriptiva es la rama de la estadística orientada a la síntesis y caracterización de datos, ya que describe las características importantes de un conjunto de datos de forma sencilla, mediante el uso de tablas, medidas y representaciones gráficas. Esto permite organizar la información e identificar patrones, tendencias y variabilidad, sin buscar probar hipótesis ni realizar inferencias estadísticas sobre una población.

En este apartado se presenta su definición y alcance general; el tratamiento metodológico detallado, así como su aplicación práctica en Excel, se desarrollan de manera sistemática en el Capítulo 4, donde se incluyen los esquemas y figuras correspondientes.

1.3. Estadística inferencial

La estadística inferencial permite obtener conclusiones poblacionales a partir de muestras, mediante la prueba de hipótesis y la generalización de resultados (Figura 1.2.).

5

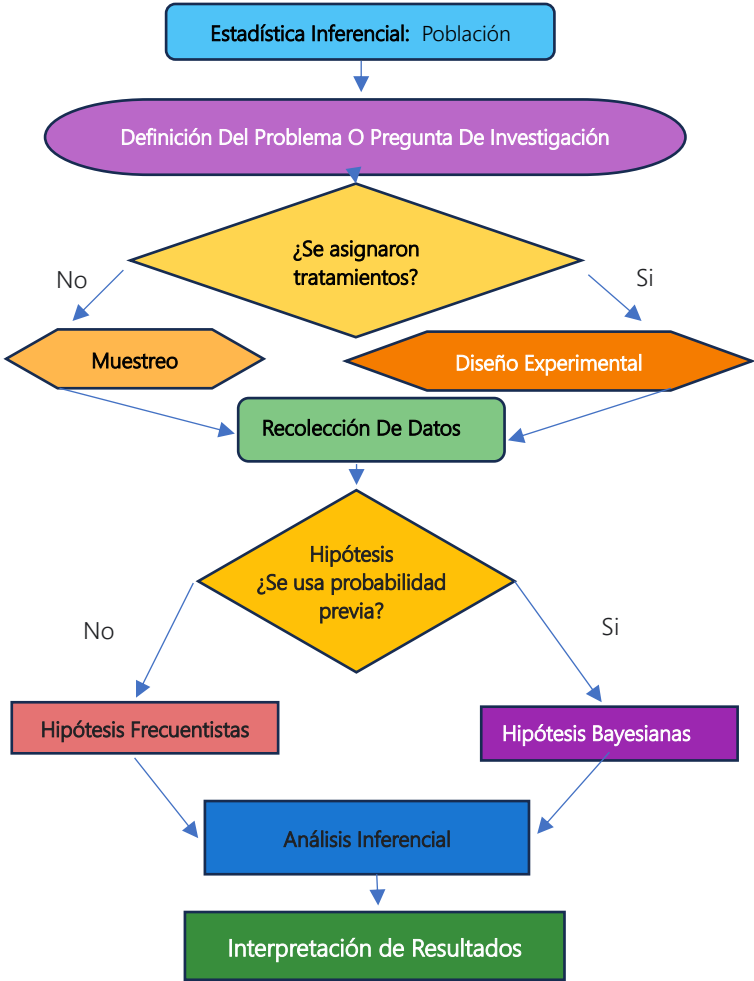


Figura 1.2. Esquema de decisión de la estadística inferencial.

1.4. Teoría del muestreo

La teoría del muestreo en Biología, Ingeniería Ambiental, permite estudiar una parte para conocer el todo (Figura 1.3.).

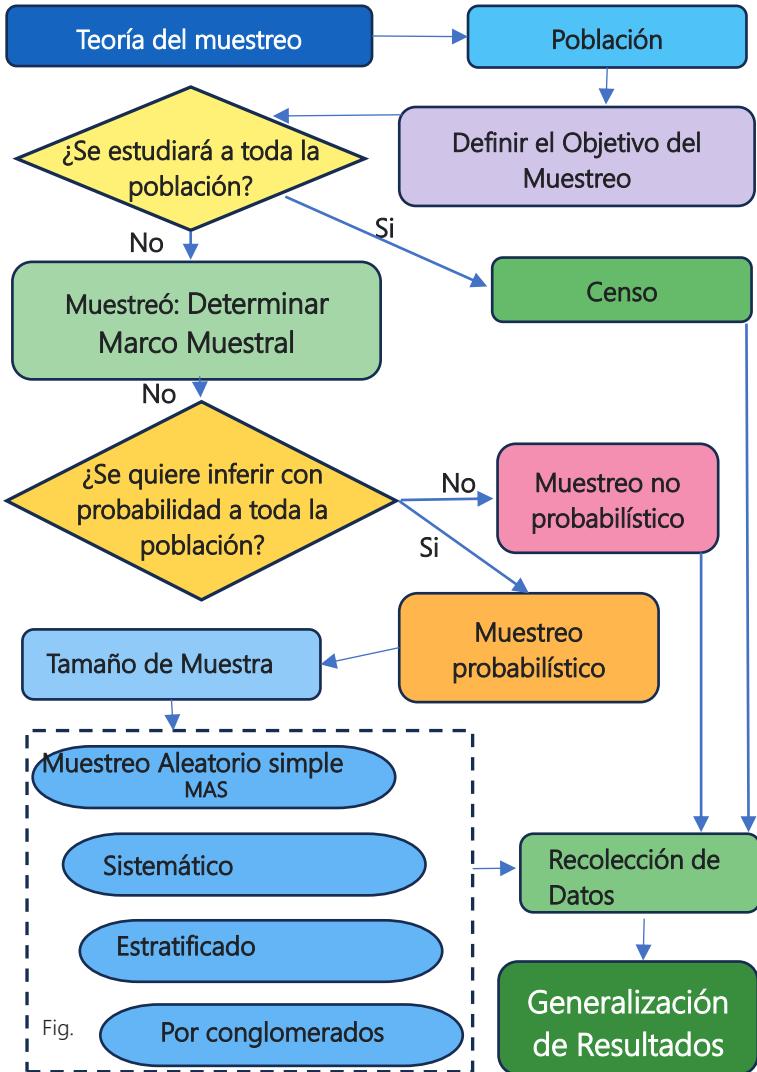


Figura 1.3. Esquema del proceso de muestreo estadístico.

1.4.1. Qué es el Muestreo

La muestra es un subconjunto representativo de la población para realizar inferencias. Según Cochran (1977), permite obtener información precisa a menor costo. Por su parte, Neyman (1934) sostiene que debe ser probabilística para garantizar su validez científica y estadística en el análisis de datos.

1.4.2. El Universo y la Población

El universo es el conjunto teórico y abstracto de todas las unidades con características comunes, sin límites de tiempo o espacio. Según Kish (1965), es un ideal inabarcable al que se busca generalizar hallazgos. En contraste, la población es el grupo finito, delimitado por lugar y tiempo, al que el investigador realmente accede; para Cochran (1977), esta debe estar claramente identificada para poder ser medida.

1.4.3. La Muestra

La muestra es un subconjunto representativo de la población, seleccionado para realizar inferencias sobre el total de la misma. Según Cochran (1977), esta selección debe permitir obtener información precisa con menor costo, mientras que Neyman (1934) fundamenta que debe ser probabilística para garantizar su validez científica y estadística.

1.4.4. Aleatorio

Lo aleatorio es el proceso donde cada elemento tiene una probabilidad determinada y no nula de ser seleccionado, eliminando el sesgo humano. Según Kish (1965) y Cochran (1977), este mecanismo garantiza que la muestra sea representativa y permite realizar inferencias estadísticas válidas.

2. Que son las variables y los datos

En estadística primero debemos conocer qué es una variable y qué es un dato.

- Variable: es la característica que queremos estudiar y que puede cambiar de un caso a otro.
Ejemplo: la edad, el peso, el contaminante, el ingreso.
- Dato: es el valor concreto que obtenemos al medir u observar la variable.
- Ejemplo: 20 años, 65 kg, 16 ppm, \$5,000.

En Excel:

- Cada columna representa una variable.
- Cada celda dentro de esa columna contiene un dato.

2.1. Elección de variable

Las variables se eligen según el objetivo del estudio. En Ingeniería Ambiental, Biología y Gestión Ambiental se seleccionan las características medibles más importantes (ejemplo: calidad del agua, biodiversidad, uso de suelo).

Se priorizan aquellas que permitan explicar, comparar o predecir fenómenos del ambiente. Para responder a tu pregunta de investigación.

2.2. Recopilación de información

La recopilación de información es el proceso de obtener los datos de las variables que elegimos. Se realiza a través de observaciones, mediciones, encuestas o registros. Es reunir datos confiables que después se analizarán en Excel.

ejercicio

2.3. Tipos de datos y su relevancia

Los datos se dividen en Variables Cualitativas y Cuantitativas. Las Cualitativas son, características o cualidades no numéricas que describen categorías, atributos o propiedades (en estadística se usan palabras y después estas se presentan con números). Y las Variables Numéricas, son características expresadas mediante números que permiten realizar mediciones y operaciones aritméticas.

Antes de los números, el orden: ¿Cómo "habla" la naturaleza con tu computadora?

El recolectar datos, a manera de ejemplo es como capturar un fragmento de la vida; sin embargo, la estadística no entiende de "intuición", solo de estructuras. Antes de correr modelos complejos, debemos decidir cómo "envasar" cada observación. Si etiquetas mal un dato desde el inicio, estarás forzando a tu software a realizar operaciones imposibles, como intentar promediar colores o sumar fechas.

La clave de un análisis exitoso no está en la fórmula final, sino en la integridad del registro previo. Ahora, daremos el salto del campo a la pantalla: elige el formato correcto de las Variables Cuantitativas:

- Enteros: Para conteos absolutos donde no existen las fracciones.
 - *Uso:* Conteo de individuos en un cuadrante o número de colonias bacterianas en una placa Petri. (No puedes tener 4.5 bacterias).
- Decimales: Para medidas de alta precisión y escalas continuas.
 - *Uso:* Registro de la biomasa de una muestra en gramos (12.45g) o la longitud del ala de un ave en milímetros.

- Porcentaje: Para expresar proporciones o éxito relativo dentro de una población.
 - *Uso:* Tasa de supervivencia de un grupo de control (82%) o cobertura de dosel en un bosque.
- Booleanos (Lógicos): La forma más eficiente de registrar presencia o ausencia (Verdadero/Falso).
 - *Uso:* ¿El ejemplar presenta el parásito? ¿La semilla germinó bajo estrés hídrico? Ideal para filtros rápidos en bases de datos.
- Texto: Para etiquetas o nombres.
 - *Ejemplo:* Nombre científico (*Escherichia coli*) o tratamiento (Control, Fertilizante A).
- Fecha/Hora: Crucial para estudios temporales.
 - *Ejemplo:* Día de siembra (15/03/2026) o duración de un ciclo (hh:mm:ss). No los escribas como texto, o no podrás calcular lapsos.
- Científico: Para números extremadamente grandes o pequeños. *Ejemplo:* Concentración de ADN (2.6×10^{-6}), es tan pequeña que se expresa en negativo el subíndice. O el conteo de esporas (3.1×10^5), es tan grande que está en positivo el subíndice.

2.4. Clasificación de datos: Escalas de Medición

El éxito de un buen análisis estadístico está en identificar bien, tu tipo de variable, usando los ejemplos según la escala de Stevens (Stevens, S. S., 1946).

1. ¿Tu dato o variable es texto o número? Si es una categoría o palabra, es categórica; Si es un número, es numérica.

2. Identifica los tres atributos de una escala de medición, el Orden, Distancia y Origen para clasificar correctamente (Figura 2.1.).

Orden: Es la capacidad de organizar los datos de menor a mayor o establecer una jerarquía (quién va primero y quién después).

Distancia: Es la existencia de un intervalo exacto, equidistante y constante entre un valor y otro, lo que permite sumar o restar los datos.

Origen: Es la presencia de un cero absoluto que indica la ausencia total de la característica medida (permitiendo multiplicar o dividir valores).

a. Los datos de tu variable ¿Tienen ORDEN?

¿Puedes organizar tus datos de menor a mayor o por jerarquía?

NO: Tus datos son Nominales (solo etiquetas o nombres).

SÍ: Pasa al siguiente punto.

b. Los datos de tu variable ¿Tienen DISTANCIA?

¿Existe un número exacto y constante entre cada valor que te permita sumar o restar?

NO: Tus datos son Ordinales (hay orden, pero no sabemos qué tanto espacio hay entre ellos o bien la distancia entre cada valor no es uniforme. (pH, Escalas de Likert, Richter,)

SÍ: Pasa al siguiente punto.

c. Los datos de tu variable ¿Tienen ORIGEN?

¿El número "cero" representa la ausencia total de lo que estás midiendo?

NO: Tus datos son de Intervalo (el cero es solo un punto de referencia).

SÍ: Tus datos son de Razón (el cero representa la ausencia de la variable y puedes comparar proporciones, como "el doble" o "la mitad").

estadística

TIPOS DE VARIABLES											
VARIABLES CATEGÓRICAS / CUALITATIVAS						VARIABLES NUMÉRICAS / CUANTITATIVAS					
NOMINAL			ORDINAL			INTERVALO			RAZÓN		
Ningún atributo			Un atributo			Dos atributos			Tres atributos		
Orden	Distancia	Origen	Orden	Distancia	Origen	Orden	Distancia	Origen	Orden	Distancia	Origen
Posee categorías a las que se asigna un nombre sin que exista ningún orden implícito entre ellas. Siendo el Orden , la capacidad de organizar los			Posee categorías ordenadas, pero no permite cuantificar la distancia entre una categoría y otra. Siendo la Distancia , la existencia de un			Tiene intervalos iguales y medibles, pero no tiene un origen real. Puede asumir valores negativos. Siendo el Origen , la presencia de un cero absoluto que			Tiene intervalos constantes entre valores, además de un origen real. El cero significa la ausencia de la variable.		
Dicotómicas			Dicotómicas			Continuas			Continuas		
									v. simple		v. compleja
Genero	Belleza	Velocidad	Volumen	Vegetación		Temperatura	Horario (hora)	Fecha	Peso	Radiación	Velocidad
Masculino	Bonito	Rápido	Vacio	Primaria	No fumadores	- 5.3 °C	0 h	1.2 semanas	1.62 kg	1.5 Sv	1.2 km/h
Femenino	Feo	Lento	Lleno	Secundaria	Fumadores	0 °C	12.5 h	3.2 semanas	5.75 kg	5.4 Sv	5.3 km/h
						5.3 °C	23.5 h	4.2 semanas	10.89 kg	10.3 Sv	10.4 km/h
Politómicas :			Politómicas :			Discretas			Discretas		
									simple		compleja
Raza	Enfermedad	Contaminantes	Educación	Contaminación	pH	Temperatura	Horario	Altitud	# de especies	# de hijos	Gravedad
Blanca	Gripe	Monóxido de CO	Primaria	Ausente	5	- 5 °C	0 h	-10 msnm	1 especie	1 hijo	1 m/s ²
Negra	Cólera	Organoclorados	Secundaria	Bajo	6	0 °C	12 h	0 msnm	5 especies	2 hijos	2 m/s ²
Amarilla	Viruela	Mercurio	Preparatoria	Moderado	7	5 °C	24 h	100 msnm	10 especies	3 hijos	3 m/s ²
Tipo de flor	Locomoción	Nitratos	Universidad	Alto	8		Ruido Db	Vulnerabilidad	# de célula	Oxígeno ppm	Radiación Bq
Alimentación	Residuo solido	Metano	Maestría	Muy alto	9		Riesgo ambiental	Eficiencia	# de bacteria	CO ₂	Tasa natalidad

Figura 2.1. Escalas de medición

2.5. Ordenamiento de datos

El ordenamiento de datos significa organizar la información que recolectamos para poder analizarla mejor. Se hace colocando los valores en una lista clara.

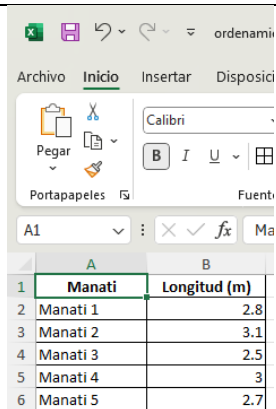
Ejemplo: Manatí de Jonuta, Tabasco (*Trichechus manatus* L.)

Un grupo de estudiantes midió la longitud (en metros) de cinco manatíes en el río Usumacinta, Jonuta, Tabasco (Figuras 2.2 y 2.3).

13

1. Datos obtenidos (sin ordenar): 2.8, 3.1, 2.5, 3.0, 2.7
2. Ordenamos los datos de menor a mayor: 2.5, 2.7, 2.8, 3.0, 3.1
3. En Excel:
 - En una columna escribimos los valores.
 - Usamos la herramienta Ordenar de menor a mayor: Excel-> Datos-> Ordenar-> Ordenar de menor a mayor
 - Así tendremos la lista para calcular promedios, gráficos y otros análisis.

Figura 2.2. Manatí



	A	B
1	Manatí	Longitud (m)
2	Manatí 1	2.8
3	Manatí 2	3.1
4	Manatí 3	2.5
5	Manatí 4	3
6	Manatí 5	2.7

Figura 2.3. Datos de longitud

3. Uso de Excel® 2016 para el Análisis de Datos Estadísticos

Este libro te permitirá trabajar en versiones de Excel® 2016, 2019, 2021 y Microsoft 365, el ambiente gráfico es similar y el manejo de sus ecuaciones es el mismo. Sus palabras clave están en Español y pueden cambiar para otros idiomas.

3.1. Que es Excel® y que es una base de datos

Excel® es un programa que permite organizar, calcular y analizar datos mediante tablas, fórmulas, gráficos, herramientas estadísticas de manera sencilla.

Una base de datos es un conjunto organizado de información, como una tabla, que permite guardar, buscar y analizar datos fácilmente.

3.2. Interfaz de Excel® 2016: celdas, filas, columnas, hojas

① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩

Al abrir el programa se abre un nuevo archivo llamado Libro ①, compuesto por hojas②.

Se presentan Pestañas con Cinta de Opciones: Archivo③, Inicio④, Insertar⑤, Disposición de página⑥, Formulas⑦, Datos⑧, Revisar⑨, Vista⑩, ayuda. En algunas versiones se pueden presentar otras cintas adicionales (Figura. 3.1.).

En cada Cinta hay un conjunto de grupos de comandos relacionados y organizados por funcionalidad. Las letras en la parte superior A, B, C, D hacen referencia a las columnas. Las filas están numeradas (1, 2, 3..) y van de arriba hacia abajo. Cada cuadro donde se cruzan se llama celda (por ejemplo, A1).

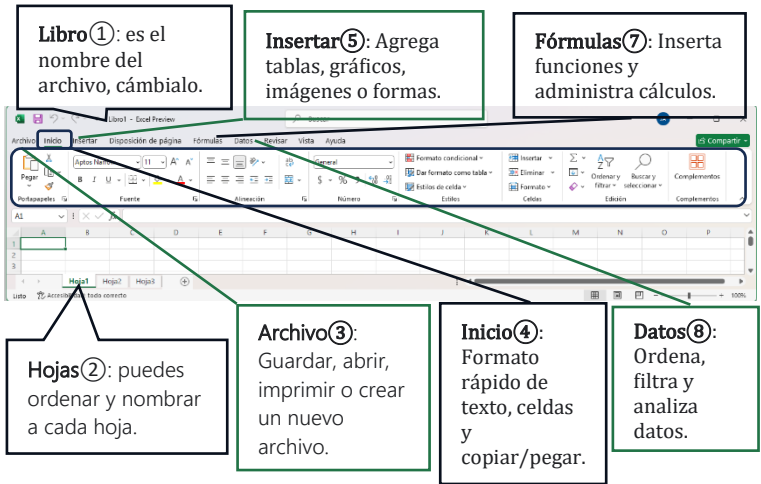


Figura. 3.1. Pestañas con Cinta de Opciones

Para dar nombre al archivo, nos vamos a la Cinta de opciones Archivo , damos clic en Guardar como①, aparece un recuadro y buscamos una carpeta donde guardar el archivo, en esta caso es en Documentos②, lo nombramos "Ejemplo.xlsx③" y finalizamos dando clic en Guardar④ (Figura. 3.2.).

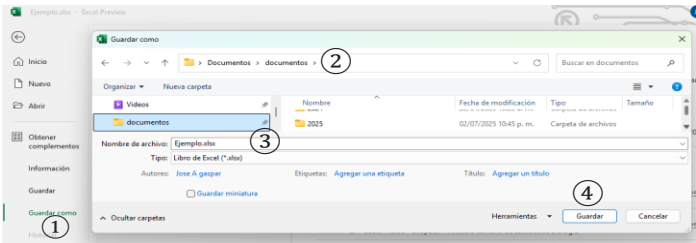


Figura. 3.2. Guardar como

3.3. Captura y formato de datos en Excel: números, texto, fechas

Para colocar Datos selecciona una celda, por ejemplo, A1①, verás su nombre en el Cuadro de nombres②.

Escribe un dato, como 1234, y observa que también aparece en la Barra de fórmulas③.

En la Cinta de opciones Inicio④, haz clic en Formato de número⑤.

Elige el tipo de dato: Número⑥, Texto, Fecha, Porcentaje, etc. (Figura. 3.3.)

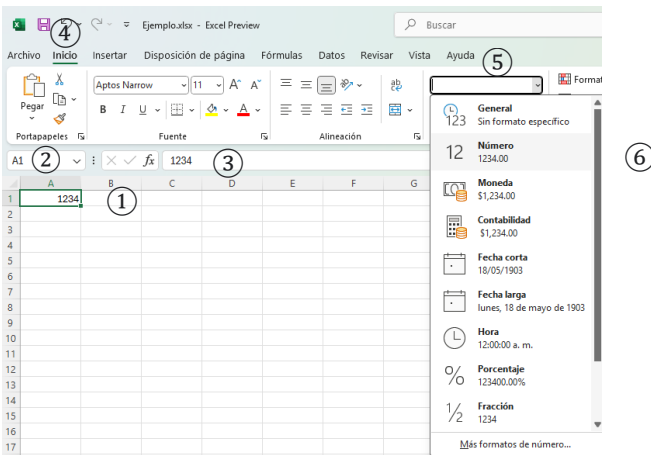


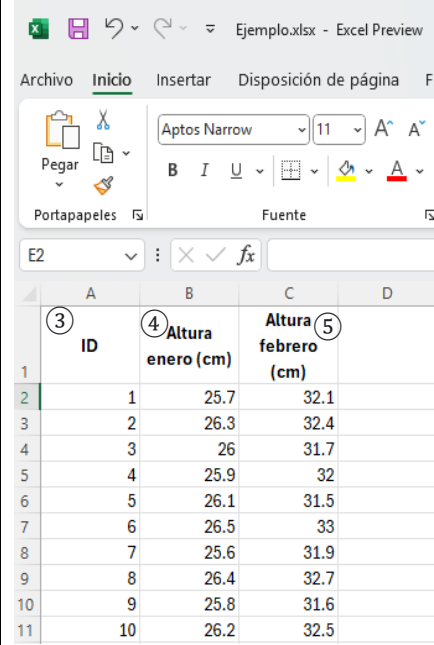
Figura 3.3. Formato de número

Nota: si le das clic derecho con el mouse a la celda te dará más opciones: explora, "Formato de celdas".

Explórala para conocer como darle diferentes formatos a la celda

3.4. Registro de datos numéricos

Registro del crecimiento mensual de 10 plantas de mangle rojo (<i>Rhizophora mangle</i> L.), la medición fue en centímetros (Figura 3.4. y 3.5.)	
	Figuras: 3.4 y 3.5.

					1 Abre Excel®.
					2 Coloca los títulos de las columnas:
					③ Haz clic en la celda A1 y escribe: ID
					④ Haz clic en la celda B1 y escribe: Altura enero (cm)
					⑤ Haz clic en la celda C1 y escribe: Altura febrero (cm)
Figura. 3.6. Datos de altura en cm					

Ingresas los datos de identificación: En la columna A, escribe los números del 1 al 10 (uno por fila), desde la celda A2 hasta A11. Esto identifica a cada planta. Continúa en las columnas B y C agregando los datos de altura (Figura. 3.6.). Importante: No dejes celdas vacías, ya que eso puede causar errores en los análisis.

3.5. Ordenar datos

Cuando tenemos datos como en la tabla anterior (Figura 4) y queremos organizar los datos sin perder el orden del número de planta. Excel® tiene más de una opción, para realizarlo veremos una forma.

1 Selecciona la tabla, haz clic izquierdo en la celda A1, mantén presionado y arrastra el cursor hasta la celda C11, luego suelta.

② Ve a la pestaña "Inicio": en Edición ③ Haz clic en el ícono de "Ordenar y filtrar" (parece un embudo con flechas).

④ Elige "Orden personalizado" (Figura. 3.7.).

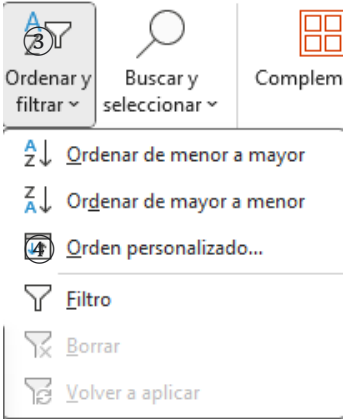


Figura. 3.7. Ordenar y filtrar

Aparece un cuadro que se llama ⑤ Ordenar; en la ventana llamada "⑥ Ordenar por", dice ID, Altura enero (cm) o Altura febrero (cm), elegimos Altura enero (cm); luego elige "⑦ Ordenar según" elegimos Valores de celda; luego "⑧ Criterio de ordenación multicriterio", te indica si deseas ordenar De menor a mayor o De mayor a menor, elige De menor a mayor; al final Presiona (también permite ordenar texto, fechas, horas, colores de celdas o fuentes, iconos, imágenes insertadas en las celdas) ⑨ Aceptar (Figura. 3.7.).

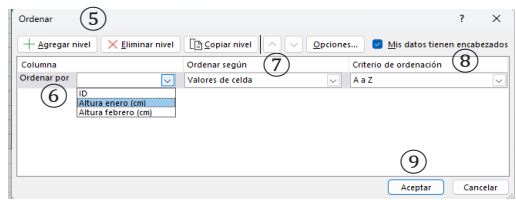
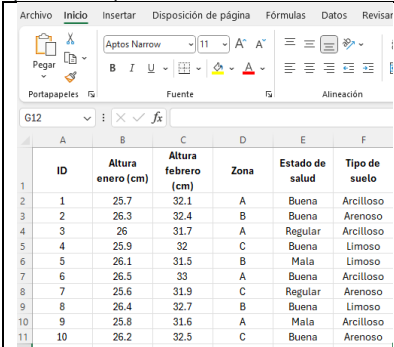


Figura. 3.7. Ordenar: los datos quedaran ordenados de menor al mayor.

3.6. Filtrar datos en Excel®

Cuando queremos ver solo ciertos datos de una tabla (Figura 3.8.), usamos el Filtrado. Sirve para mostrar solo la información que cumple con una condición.



	A	B	C	D	E	F
	ID	Altura enero (cm)	Altura febrero (cm)	Zona	Estado de salud	Tipo de suelo
1	1	25.7	32.1	A	Buena	Arcilloso
2	2	26.3	32.4	B	Buena	Arenoso
3	3	26	31.7	A	Regular	Arcilloso
4	4	25.9	32	C	Buena	Limoso
5	5	26.1	31.5	B	Mala	Limoso
6	6	26.5	33	A	Buena	Arcilloso
7	7	25.6	31.9	C	Regular	Arenoso
8	8	26.4	32.7	B	Buena	Limoso
9	9	25.8	31.6	A	Mala	Arcilloso
10	10	26.2	32.5	C	Buena	Arenoso

Figura 3.8. Datos de salud

De la tabla de 10 plantas de mangle rojo (*Rhizophora mangle* L.), tenemos 5 variables, queremos saber de la variable "Estado de salud" cuales están sanas (Figura. 3.8.). Selecciona con el mouse, la tabla completa, incluyendo los encabezados (por ejemplo, desde A1 hasta F11).

19

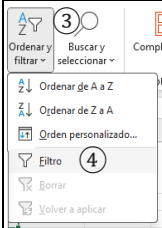
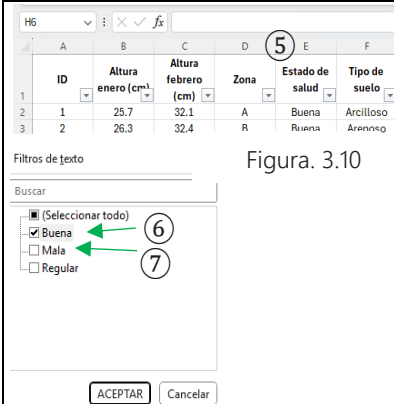


Figura 3.9. Ordenar y filtrar

Ve a la Cinta "①Inicio", luego ②Edición, luego "③Ordenar y filtrar", después haz clic en el botón "④Filtro" (ícono de embudo) (Figura 3.9.).



	A	B	C	D	E	F
	ID	Altura enero (cm)	Altura febrero (cm)	Zona	Estado de salud	Tipo de suelo
1	1	25.7	32.1	A	Buena	Arcilloso
2	2	26.3	32.4	B	Buena	Arenoso

Filtros de texto

Buscar

☒ (Seleccionar todo)

☒ Buena ⑥

☐ Mala

☐ Regular

ACEPTAR ⑦ Cancelar

Figura 3.10

Aparecerán flechas pequeñas en los encabezados de cada columna. Da clic en la flecha del encabezado "⑤Estado de salud" (Fig. 9). Luego marca la casilla ⑥Buena y dale ⑦ACEPTAR (Figura. 3.10. y 3.11.).

Figura 3.11. Filtros de texto

Dejará solo las Buenas y ocultará las demás.

3.7. Operaciones básicas en Excel®: suma, resta

En el Excel® podemos realizar operaciones desde una celda. Para ello iniciamos una fórmula con el símbolo “=” o el “+”. Es muy sencillo, veremos una suma como ejemplo

①Da clic en la celda A1 (en el Cuadro de nombre② puedes ver el código de la celda).

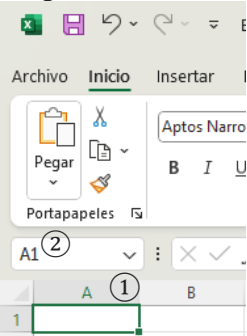


Figura. 3.12. Cuadro de nombre

③Escribe la siguiente fórmula aritmética con el teclado: =23+12 (Fig. 12).

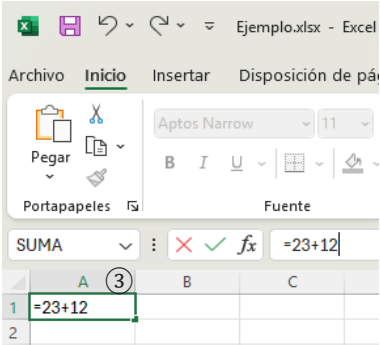


Figura. 3.13. Realizando una suma

Presiona la tecla Enter ↵, Listo, Excel® mostrará el resultado④: 35 (Figura. 3.14).

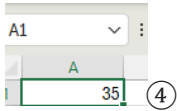


Figura. 3.14. Resultado

Cuadro 1: En Excel® puedes realizar diversas operaciones

Operación	Símbolo en Excel®	Ejemplo (escribe en una celda)	Resultado
Suma	+	=23+12	35
Resta	-	=45-15	30
Multiplicación	*	=6*7	42
División	/	=50/5	10
Potencia	^	=5^3	125
Raíz cuadrada	RAIZ ()	=RAIZ (36)	6
Logaritmo base 10	LOG10 ()	=LOG10 (1000)	3
Logaritmo natural	LN ()	=LN (2.71828)	1
Aritmética compleja	()	=(5+3)*2	16
Nota: El nombre las funciones cambia en otros idiomas			

3.8. Uso de funciones estadísticas en Excel®

En Excel puedes tu elaborar la fórmula o bien utilizar formulas ya instaladas en el programa, como en la Barra de fórmulas.

El recuadro de Insertar función aparece al hacer clic en el ícono ***fx*** que se encuentra a la izquierda de la barra de fórmulas.

Sirve para buscar y usar funciones ya elaboradas por Excel, sin necesidad de escribirlas manualmente, las cuales son 484 en total.

Damos clic en seleccionar una categoría (hay 14 categorías).

Seleccionamos la categoría de: Estadísticas, y le damos en Aceptar (Figura. 3.15.).

Figura. 3.16. Insertar función: Estadísticas

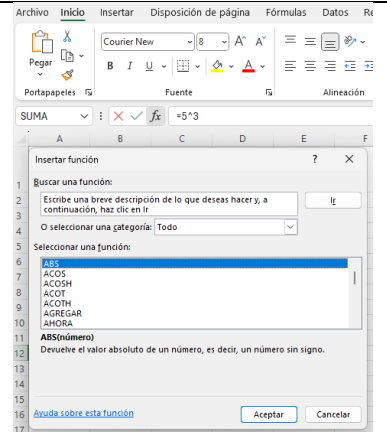
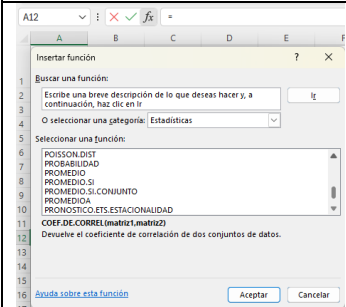


Figura. 3.15. Insertar función: Todo



Nos presenta una lista de 90 funciones de Estadística en orden alfabético.

La función Promedio es una de las más usadas en estadística.

Buscamos la función de Promedio y le damos Aceptar (Figura. 3. 16.).

3.9. Uso de Funciones

En matemáticas, una ecuación tiene un signo igual y busca una solución (por ejemplo, $2X + 5 = 7$).

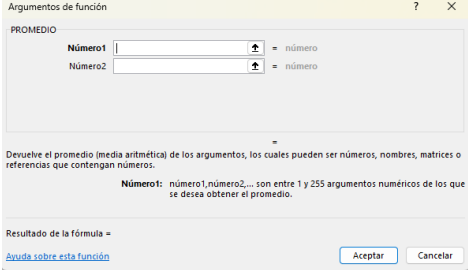
En Excel, lo que escribes en una celda (como =PROMEDIO(A1:A5)) es llamado función, porque realiza el cálculo con datos de un cuadro.

1. Nombre de la función: aparece arriba del recuadro; por ejemplo: PROMEDIO.

2. Casillas para escribir datos: aquí escribes los números o rangos de celdas que quieres usar, como B2:B11. Lo escribes sin espacios o seleccionas con el mouse.

3. Descripción de la función: Debajo de las casillas, te dice qué hace la función. Ejemplo: "Devuelve el promedio..."

4. Resultado automático: en la esquina inferior izquierda verás el resultado que Excel calcula con lo que escribiste (Figura 3.17.).



5. Botones finales: Aceptar, pone el resultado en la celda; Cancelar, cierra el cuadro sin guardar nada.

Figura. 3.17. Argumento de función: promedio

Promedio de la Altura enero (cm)

1 Celda A12

2 Clic en **fx** : Categoría estadísticas: Promedio: Aceptar

3 Número 1: selecciona las celdas con los datos numéricos de las 10 plantas (B2:B11).

4. Resultado de la Fórmula 26.05.

5. ACEPTAR (Figura. 3.18).



Figura. 3. 18

4. Estadística descriptiva

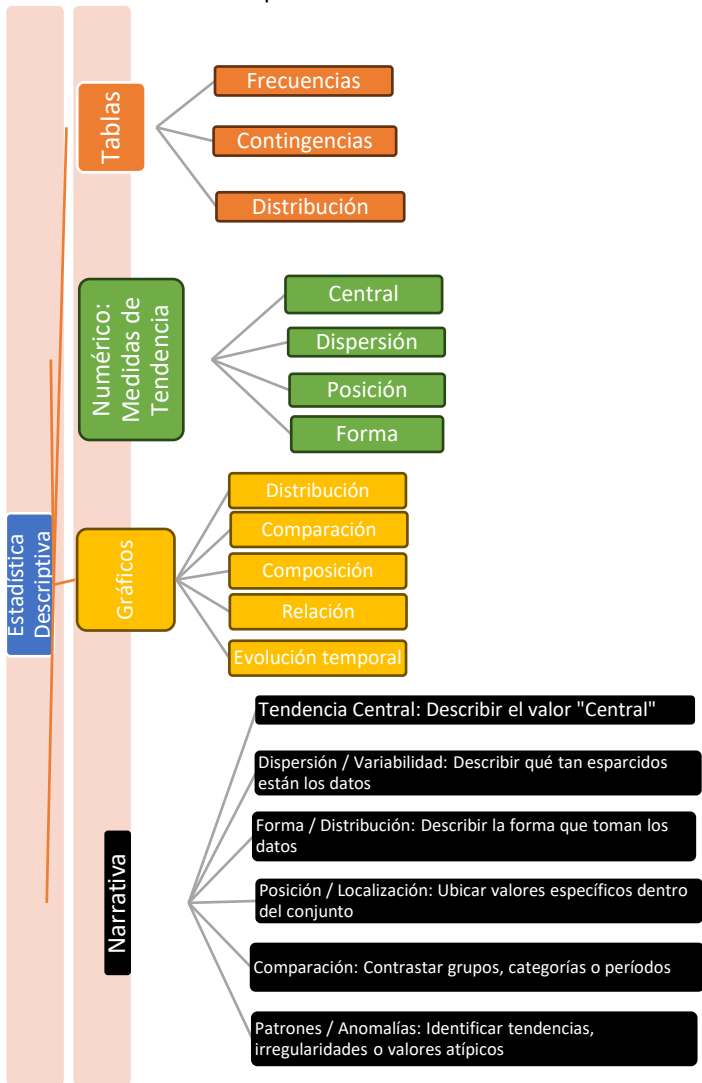


Figura. 4.1. Estadística Descriptiva

4.1. Estadística descriptiva numérica

Es la rama que resume y caracteriza datos numéricos sin inferir sobre la población; describe “cómo son” tus datos con medidas (no son pruebas).

Se realiza con la estimación de parámetros que es el proceso de inferir un valor desconocido (parámetro) de una población utilizando datos de una muestra (por ejemplo, la media o la proporción).Y se hace con fórmulas específicas según el parámetro que quieras calcular.

Grupo	Medidas comunes (función de Excel entre paréntesis)
Tendencia central	Media; Mediana; Moda
Dispersión	Desviación estándar (muestral); Desviación estándar (poblacional); Varianza (muestral); Varianza (poblacional) ; Rango RIC / IQR ; Coeficiente de variación (CV); Desviación absoluta mediana
Forma	Asimetría; Curtosis
Posición	Mínimo; Q1; Mediana; Q3; Máximo; Percentil p

En Excel se encuentra diferentes tipos de medidas de Medias, Promedios, las cuales tiene diversas aplicaciones en Ingeniería ambiental, Biología y Gestión Ambiental

Medida (función de Excel)	Usos en ingeniería ambiental	Usos en biología	Usos en gestión ambiental
Media aritmética (PROMEDIO)	Concentración promedio de contaminantes en agua/aire.	Peso/longitud promedio de organismos.	Indicadores promedio (residuos, consumo de energía).
Mediana (MEDIANA)	Turbidez típica evitando atípicos.	Tamaño de semilla/talla típica robusta.	Reporte "valor típico" resistente a outliers.
Moda (MODA.UNO)	Tamaño de partícula más frecuente.	Talla modal de peces.	Categoría/valor más frecuente en monitoreo.
Media geométrica (MEDIA.GEOM)	Promedio de concentraciones log-normales (coliformes).	Tasas de crecimiento poblacional/composición química.	Índices compuestos multiplicativos.
Media armónica (MEDIA.ARMO)	Promediar tasas (velocidad de flujo, caudal específico).	peces/hora, individuos por trampa-noche, aves/hrs de observación	Eficiencias promedio (litros/kWh, t/km).
Media acotada (MEDIA.ACOTADA)	Promedio sin extremos en series ruidosas.	Medidas morfométricas con valores atípicos.	Indicadores robustos para auditorías.
Media condicional (PROMEDIO.SI / PROMEDIO.SI.CONJUNTO)	Promedio por sitio/temporada que cumple norma.	Promedio por especie/sexo/edad.	Promedio por programa/área/responsable.
Desviación estándar muestral (DESVEST.M)	Variabilidad de PM2.5/DBO.	Variación de talla/peso.	Estabilidad de indicadores (control).
Varianza muestral (VAR.S)	Comparar dispersión entre estaciones.	Variabilidad genética/morfológica.	Comparar variabilidad entre proyectos.
Coefficiente de variación (DESVEST.M/PROMEDIO)	Homogeneidad de procesos (tratamiento de agua).	Variación relativa entre poblaciones.	Comparabilidad de indicadores (%).
Rango (MAX-MIN)	Amplitud de pH/OD en un día.	Amplitud de tamaños/biomasa.	Variación del indicador clave entre periodos
RIC / IQR	Dispersión robusta de DQO/NO ₃ ⁻ .	Dispersión de tallas sin extremos.	Variabilidad central de costos/impactos.
(CUARTIL.INC(...,3)-CUARTIL.INC(...,1))	Sesgo en metales pesados/coliformes.	Sesgo en distribución de tallas.	Detección de colas en datos de gestión.
Asimetría (SESGO)	Colas pesadas en contaminantes.	Pico/colas en rasgos morfológicos.	Riesgo de eventos extremos en métricas.
Curtosis (CURTOSIS)	P95 de NO ₃ , PM10 para norma.	p90 de talla de madurez.	Percentiles para metas/umbrales.
Percentil p (PERCENTIL.INC)	Rangos por sitio/temporada.	Cuartiles de biomasa/talla.	Resumen por cuartiles en tableros.
Cuartiles Q1-Q3 (CUARTIL.INC)	Picos y mínimos de caudal, OD, pH.	Máx/min de abundancia o talla.	Valores extremos en indicadores clave.
Mínimo / Máximo (MIN / MAX)	Posición de una estación vs otras.	Ranking de individuos por medida.	Escala/fon de desempeño de áreas.
Jerarquía (JERARQUIA.MEDIA)	Serie con cerros/lógicos incluidos.	Registros con VERADERO/FALSO (presencia).	Encuestas con valores lógicos.
Promedio con texto (PROMEDIOA)	Promedio filtrado por criterios (DAVERAGE).	Promedio por atributos en fichas.	Promedio por campos en bases ambientales.
Promedio de base de datos (BDPROMEDIO)			

Figura 4.2. Medidas descriptivas en excel

Estadística aplicada a las ciencias ambientales y de la vida: una perspectiva computacional con Excel.

4.2. Cómo elegir el tipo de Gráfico

Cómo usar el Árbol de Decisiones de Gráficos en Excel

Para elegir el gráfico adecuado, se utiliza el Árbol de Decisiones de Gráficos en Excel, el cual permite seleccionar de manera ordenada el tipo de gráfico según el tipo de datos, su escala de medición y el objetivo del análisis.

1. Mira tus datos. Pregunta: ¿Son nombres, rangos, números o mezclas?
2. Identifica y elige el tipo de escala:
Nominal, nombres o categorías (ej. especie).
Ordinal, con orden o rangos (ej. calidad del agua).
Intervalo/Razón, números (ej. precipitación, precios).
Escala combinada, meses de lluvia y temperatura).
3. Busca el ejemplo parecido a tu caso. Ejemplo: "Peso vs talla de manatíes" → usa Dispersión (XY).
4. Elige el gráfico indicado. Solo copia tus datos en Excel y selecciona ese gráfico (Fig.).

27

Árbol de Decisiones para Elegir Gráfico en Excel 2016

1. ¿Qué tipo de datos tienes?

→ A. Nombres o categorías (Nominal)

- ◇ Número de peces por especie → Columnas o Barras
- ◇ Porcentaje de cultivos por tipo → Circular
- ◇ Casos de contaminación por estados → Mapa
- ◇ Clasificación de plantas medicinales (familia → género

→ especie) → De rectángulos o Proyección solar

→ B. Órdenes o rangos (Ordinal)

◇ Calidad del agua en tres lagunas → Columnas agrupadas o Barras agrupadas

◇ Etapas de filtración de agua y pérdida en cada paso
→ Embudo

→ C. Números en el tiempo (Intervalo / Razón – 1 variable o series temporales)

- ◊ Lluvia mensual en Tabasco → Líneas
- ◊ Agua acumulada por meses → Áreas
- ◊ Ganancias y gastos de una cooperativa → Cascada
- ◊ Precio diario del cacao → Cotizaciones

→ D. Comparación entre muchos números (Intervalo / Razón – bivariables o multivariantes)

- ◊ Peso vs talla de manatíes (*Trichechus manatus* L) →

Dispersión (X Y)

- ◊ Relación → Superficie
- ◊ Perfil de nutrientes en el suelo → Radial
- ◊ Frecuencia de tamaños de semillas → Histograma
- ◊ Distribución de altura de mangles → Cajas y bigotes

(Box Plot)

→ E. Quieres ver varias cosas a la vez (Combinado)

- ◊ Lluvia con columnas y temperatura con líneas →

Combinado

Forma práctica rápida:

- Nominal → Columnas, Barras, Circular, Mapa, De rectángulos, Proyección solar
- Ordinal → Columnas ordenadas, Barras ordenadas, Embudo
- Intervalo → Líneas, Áreas, Dispersión, Histograma, Cajas y bigotes, Radar, Superficie
- Razón → Cotizaciones, Cascada, Líneas, Áreas, Dispersión, Superficie
- Combinado → Combo Chart

4.2.1. Gráfico de Columnas

Un biólogo requiere identificar las especies de peces más frecuentes en el río Grijalva, que atraviesa el estado de Tabasco. En el año 2026 se llevó a cabo un muestreo ictiológico en diferentes puntos de los ríos de Tabasco, México, se registraron ocho especies (Cuadro:).

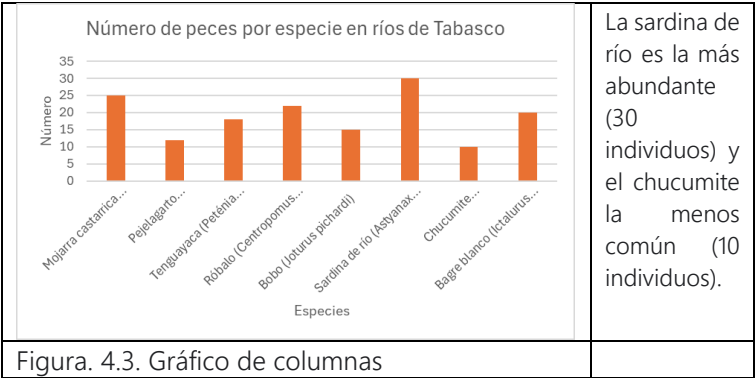
	A	B
	Especie (Nombre común / Científico)	Número de individuos
1		
2	Mojarra castarrica (<i>Cichlasoma urophthalmus</i>)	25
3	Pejelagarto (<i>Atractosteus tropicus</i>)	12
4	Tenguayaca (<i>Peténia splendida</i>)	18
5	Róbalo (<i>Centropomus undecimalis</i>)	22
6	Bobo (<i>Joturus pichardi</i>)	15
7	Sardina de río (<i>Astyanax aeneus</i>)	30
8	Chucumite (<i>Centropomus parallelus</i>)	10
9	Bagre blanco (<i>Ictalurus meridionalis</i>)	20

Cuadro:

Pasos en Excel

Paso 1 → Abrir Excel > Insertar tabla de datos en columnas: Columna A = Especie (nombre común y científico); Columna B = Número de individuos > Seleccionar rango A1:B9 > → Ir a menú → Insertar → Gráfico de Columnas → Columnas agrupadas

Ajustar el título del gráfico → "Número de peces por especie en ríos de Tabasco" > Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X): Horizontal secundario (Eje Y) los números (Figura. 4.3.).



4.2.2. Gráfico de Líneas

En 2024, la CONAGUA registró la precipitación mensual en H. Cárdenas, Tabasco mediante estaciones meteorológicas locales. Estos datos permiten conocer el comportamiento de las lluvias a lo largo del año para apoyar la planeación agrícola, la gestión del agua y la prevención de inundaciones.

Abrir Excel > Nueva hoja > Insertar datos de la tabla en columnas A y B > Seleccionar rango A1:B13

Ir a Menú "Insertar" > Gráficos > Línea > "Línea con marcadores" > Ajustar título del gráfico → escribir "Lluvia mensual en H. Cárdenas, Tabasco (mm)"

	A	B
1	Mes	Precipitación (mm)
2	Enero	121
3	Febrero	98
4	Marzo	54
5	Abril	48
6	Mayo	105
7	Junio	210
8	Julio	200
9	Agosto	249
10	Septiembre	344
11	Octubre	307
12	Noviembre	193
13	Diciembre	136
14	Fuente: CONAGUA, 2024	

Cuadro.

Dale Clic dentro de la Gráfica> Diseño de gráfico(arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X → colocar Meses; Eje Y → Precipitación (mm))> Agregar fuente (CONAGUA, 2024) debajo > Guardar archivo

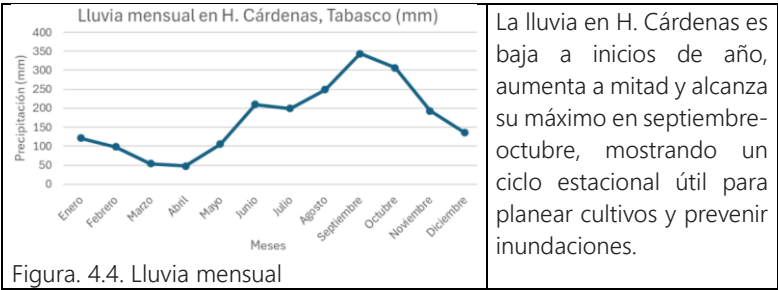
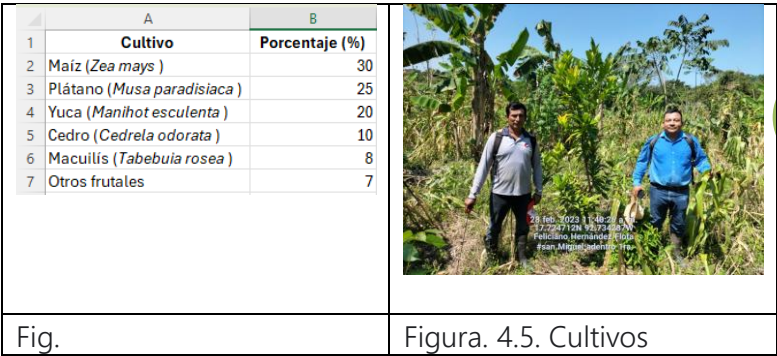


Figura. 4.4. Lluvia mensual

4.2.3. Gráfico Circular

Porcentaje de cultivos por tipo en parcelas campesinas con el cultivo de maíz en el 2025, en San Miguel Adentro, Jalapa, Tabasco.



Abrir Excel > Insertar tabla con dos columnas: A (Cultivo), B (Porcentaje)> Seleccionar rango A1:B7 > → Ir a "Insertar > Gráficos" > Elegir "Circular"

Ajustar título → "Porcentaje de cultivos en San Miguel Adentro, Jalapa, Tabasco" >

Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Etiquetas de datos> Extremo externo: Leyenda > Derecha > Verificar colores distintos por cultivo> → Guardar archivo

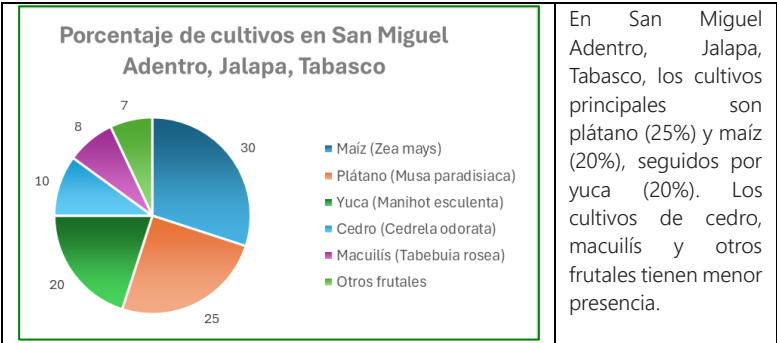


Figura. 4.6. Porcentajes

4.2.4. Gráfico de Barras

Número de peces marinos por especie (nombre común y científico) registrados en el puerto de Sánchez Magallanes, H. Cárdenas, Tabasco, a partir del conteo de ejemplares capturados.

A		B
1	Especie	Número de individuos
2	Robalo Lubina (Centropomus undecimalis)	35
3	Jurel (Caranx hippos)	28
4	Sierra (Scomberomorus maculatus)	22
5	Mojarra rayada (Eugerres plumieri)	40
6	Bagre (Ariopsis felis)	30



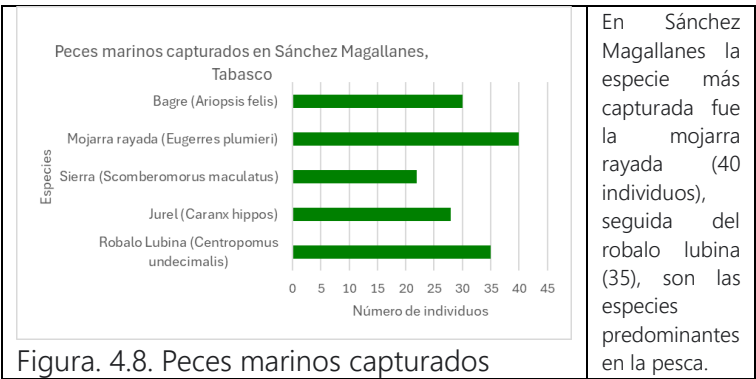
Cuadro.	Figura. 4.7. Robalo
---------	---------------------

Abrir Excel > Insertar tabla con dos columnas: A (Especie), B (Número)> Seleccionar rango A1:B6> → Ir a "Insertar > Gráficos" > Elegir "Gráfico de Barras".

→ Ajustar título → " Peces marinos capturados en Sánchez Magallanes, Tabasco"> Cambiar orientación de las barras si es necesario (horizontal o vertical)

Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X: Número de individuos); Horizontal secundario (Eje Y: Especies)

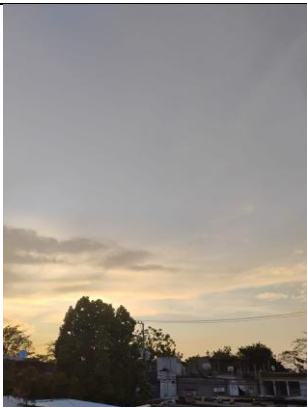
→ Guardar archivo.



4.2.5. Gráfica de Áreas

Acumulación mensual de lluvia en Tabasco, basada en datos promedio estatal 1981–2010. El agua acumulada anual es la suma de la lluvia de cada mes con la de los meses anteriores.

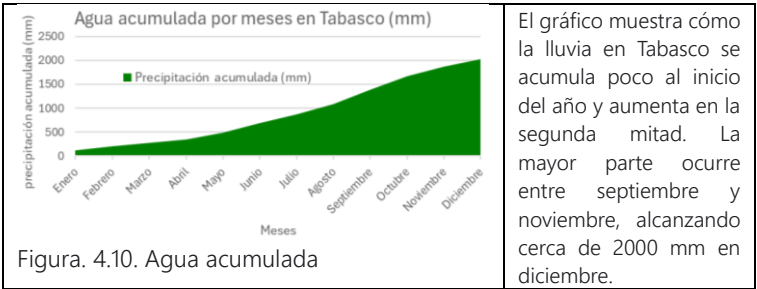
	A	B	C
	Mes	Lluvia mensual (mm)	Precipitación acumulada (mm)
1			
2	Enero	120	120
3	Febrero	90	210
4	Marzo	60	270
5	Abril	70	340
6	Mayo	150	490
7	Junio	200	690
8	Julio	180	870
9	Agosto	220	1090
10	Septiembre	300	1390
11	Octubre	280	1670
12	Noviembre	200	1870
13	Diciembre	150	2020



Cuadro.

Figura. 4.9. Nubes

Abrir Excel > Insertar tabla con tres columnas: A (Mes), B (Lluvia mensual), C (Acumulada)> → Calcular acumulada: C2 = B2; C3 = C2 + B3; copiar hasta C13
→ Seleccionar rango A1:C13> → Insertar > Gráficos > Áreas > Gráfico de áreas apiladas (solo acumulada visible: Filtro de gráficos a la derecha del gráfico, desmarcar lluvia mensual, aplicar) > → Título: "Agua acumulada por meses en Tabasco (mm)":
→ Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X: Meses); Horizontal secundario (Eje Y: Precipitación acumulada (mm)) > Leyenda: mostrar "Lluvia acumulada (mm)"> → Guardar archivo



4.2.6. Gráfica X Y (dispersión)

Un equipo monitorea manatíes (*Trichechus manatus* L.) para evaluar cómo cambia el peso (kg) con la talla lineal (m) y si la relación es aproximadamente alométrica. En la literatura con manatíes de Florida se reportan longitudes 1.47–3.23 m y pesos 77–751 kg, con crecimiento de peso cercano a una potencia de la talla (L^b); L es Longitud y b es exponente alométrico el cual se va a calcular. Use un conjunto pequeño de individuos para comparar muchos valores (escala razón, bivariado) y estimar la relación.

A1				Manatí
	A	B	C	
	Manatí	Longitud (m), X	Peso (kg), Y	
1				
2	Manatí 1	1.55	85	
3	Manatí 2	1.7	105	
4	Manatí 3	1.85	130	
5	Manatí 4	2	170	
6	Manatí 5	2.2	230	
7	Manatí 6	2.4	300	
8	Manatí 7	2.6	380	
9	Manatí 8	2.8	460	
10	Manatí 9	3	560	
11	Manatí 10	3.1	620	
12	Manatí 11	3.2	700	
13	Manatí 12	3.23	740	
14	Fuente: Harshaw, L. T., et al., 2016			

Cuadro.



Figura. 4.11. Manatíes

En: A1= Manatí; B1= Longitud (m); C1=Peso (kg); colocar datos: seleccionar de la A1:C13> Insertar gráfico: X Y dispersión: dispersión con líneas rectas y marcadores>

Los puntos muestran que, a mayor longitud, el peso aumenta y se acelera (curva ascendente). La tendencia potencia $W=22.055L^{(2.965)}$ describe casi todo el patrón ($R^2=0.998$), Figura.

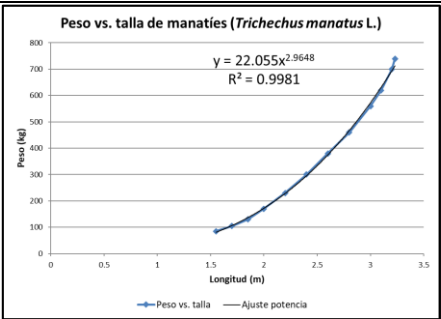


Figura. 4.12. Peso y talla de manatíes

4.2.7. Gráfica de Mapa

La calidad del aire (ICA+ US) en 5 estados de la República Mexicana y sus valores promedio de Índice de Calidad del Aire (ICA+ US) en 2024 (valores aproximados basados en IQAir) (IQAir, 2025).

	A	B	
1	Entidad	ICA+ US Promedio	
2	Ciudad de México	36	
3	Aguascalientes	28	
4	Baja California	53	
5	Tamaulipas	38	
6	Veracruz	44	



Cuadro.

Figura. 4.13. Mechero

Abrir Excel > Crear nueva hoja > → Ingresar tabla de datos en rango A1:B6 (incluye encabezados)> Seleccionar rango A1:B6> → Ir a Insertar > Gráficos > Mapas > Mapa con regiones rellenas> → Si Excel no reconoce el estado, editar las celdas con "Estado, México"> Ej.: "Ciudad de México, México"; "Veracruz, México" >→ Ajustar título del gráfico: "Contaminación del aire por estado – México (2024)". → Formato de serie de datos > Escala de colores (verde-amarillo-rojo según ICA)> → Mostrar etiquetas de datos (valores ICA) y activar leyenda> → Añadir pie de fuente: "Fuente: IQAir, 2025"



Figura. 4.14. Contaminación del aire por estados

Este mapa nos dice qué estados tienen más o menos contaminación del aire: Baja California y Aguascalientes presentan mejor calidad del aire, mientras que los demás se ubican en niveles intermedios. Los colores y los números hacen más fácil entender y comparar.

4.2.8. Gráfica de Cotizaciones

Un exportador de cacao mexicano quiere evaluar la volatilidad de precios internacionales antes de cerrar un contrato de venta a futuro. Para ello, revisa el comportamiento del contrato US Cocoa Futures (ICE) durante junio y julio de 2025. Con los datos diarios de apertura, máximo, mínimo y cierre, el exportador utilizará un gráfico de cotización para identificar: los días con mayor diferencia entre precio máximo y mínimo. Esto le permitirá decidir si conviene vender ahora o esperar, minimizando riesgos del mercado.

G8						Abrir Excel 2016> Copiar los datos de la tabla anterior en el rango A1:E8 (con encabezados en la fila 1)>
	A	B	C	D	E	
1	Fecha	Apertura	Máximo	Mínimo	Cierre	
2	17-jul-25	7500	7500	7208	7309	
3	18-jul-25	7444	7835	7400	7800	
4	21-jul-25	7750	8234	7708	8156	
5	22-jul-25	8074	8247	8022	8149	
6	24-jul-25	8440	8570	8066	8124	
7	25-jul-25	8145	8363	8093	8329	
8	31-jul-25	8248	8552	8184	8506	
9	Fuente: Investing.com, 2025					

Seleccionar el rango A1:E8> Ir a la pestaña "Insertar"> En el grupo "Gráficos", seleccionar "Todos los gráficos"> Elegir la categoría "Cotizaciones"> Seleccionar el tipo "Apertura, máximos, mínimos, cierre"> Hacer clic en "Aceptar" para insertar el gráfico> Agregar título: "Fechas seleccionadas de negociación en julio de 2025"

Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X: Fecha); Horizontal secundario (Eje Y: USD por tonelada métrica) → clic derecho sobre una figura oscura> relleno: color rojo; figura blanca > relleno: color verde)



Figura. 4.15 Gráfica de cotizaciones.

4.2.9. Gráfica de Superficie

Un biólogo requiere estimar la biomasa de sargazo (*Sargassum natans* L.y *Sargassum fluitans* L.) en las playas del municipio de Centla, Tabasco. Para ello, en el año 2026 se llevó a cabo un muestreo cuadrático de 4 × 4 metros, registrando la biomasa en gramos por cada metro cuadrado en distintos puntos de la playa (Cuadro:).

	A	B	C	D	E
1		1	2	3	4
2	A	352	685	520	356
3	B	121	170	271	716
4	C	264	180	308	337
5	D	622	349	380	558



Figura 4.16. Retícula

Abrir Excel y copiar los datos en una tabla (4 filas × 4 columnas).

Seleccionar el rango completo: de la A1:E5

Ir a Menú: Insertar :Gráfico de Superficie :Superficie 3D.

Ajustar el título del gráfico: "Distribución de biomasa macroalgal (g). Personalizar ejes: Eje X (Columnas); Eje Y (gramos de biomasa); Eje Z (Filas): gramos de biomasa.

La barra de color sirve para visualizar la concentración de masa algal debajo del agua. Los tonos verdes indican baja biomasa (0–200 g), mientras que los colores púrpura, azul y verde azulado representan aumentos progresivos hasta 800 g.

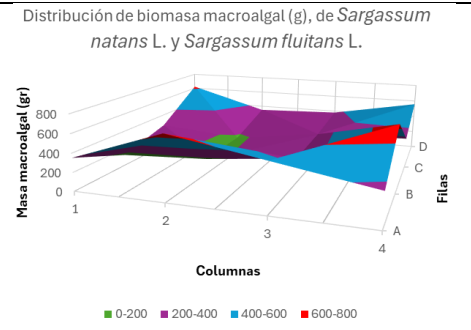


Figura. 4.17. Gráfica de superficies

La mayor concentración se observa en los colores rojo y azul, lo que sugiere mayor acumulación de *Sargassum spp* (Figura.4.17.).

B10	▼	✕	✓	f_x	
	A	B	C	D	
1	Estado	N (mg/kg)	P (mg/kg)	K (mg/kg)	
2	Chiapas	1200	18	170	
3	Veracruz	1050	22	160	
4	Jalisco	900	26	185	
5	Hidalgo	800	20	150	
6	Yucatán	700	24	140	
7	Fuente: López-Mata, L. 2022.				

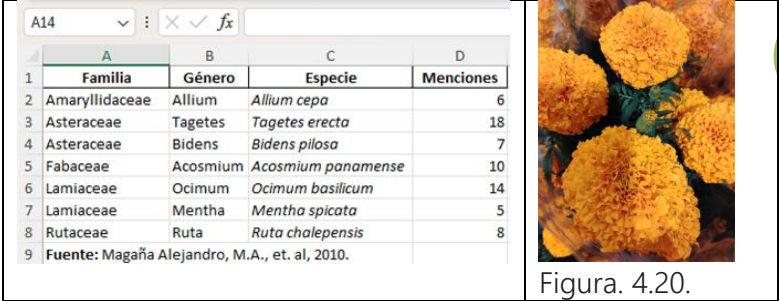
Perfil de nutrientes del suelo (N, P, K) — 5 estados de la República Mexicana

Estado	N (mg/kg)	P (mg/kg)	K (mg/kg)
Chiapas	1100	200	250
Veracruz	800	300	250
Jalisco	700	200	250
Hidalgo	500	200	250
Yucatán	400	400	250

El gráfico muestra cómo varían tres nutrientes del suelo, en cinco estados de México. Cuanto más alejada esté la línea del centro, mayor es la cantidad de ese nutriente.

4.2.11. Gráfica de rectángulos

Un Gestor Ambiental realiza un estudio en una comunidad de Nacajuca, Tabasco, para identificar el uso tradicional de plantas medicinales, organizadas por familia, género y especie (Fig.), quiere saber cuáles son más empleadas, para realizar un programa de manejo sustentable. Usará un gráfico de rectángulos que permite comparar mejor las proporciones.



Abrir Excel 2016> Ingresar la tabla en A1:D6 (encabezados y datos)> Seleccionar todo el rango A1:D6> → Ir a Insertar > Insertar gráfico De rectángulos > Ajustar el título del gráfico: "Uso tradicional de plantas medicinales en Nacajuca, Tabasco" > Mostrar etiquetas con nombre de especie y valor de menciones > Aplicar colores distintos por Familia para claridad visual

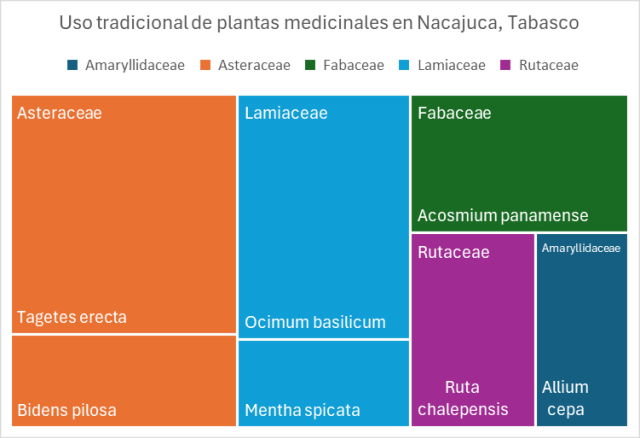
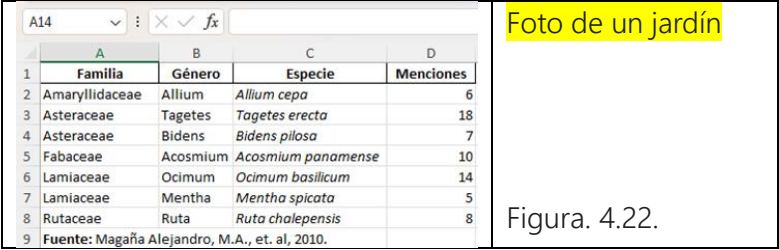


Figura. 4.21. Gráfica de rectángulos

En cuanto más grande el rectángulo, mayor es el número de menciones. Aquí, la familia Asteraceae con la especie *Tagetes erecta* es la más frecuente (Figura. 4.21).

4.2.12. Gráfica de Proyección solar

Un Gestor Ambiental realiza un estudio en una comunidad de Nacajuca, Tabasco, para identificar el uso tradicional de plantas medicinales, organizadas por familia, género y especie (Fig.), quiere saber cuáles son más empleadas, para realizar un programa de manejo sustentable. Usará un gráfico de proyección solar para mostrar las jerarquías de los datos.



Abrir Excel 2016> Ingresar la tabla en A1:D6 (encabezados y datos)> Seleccionar todo el rango A1:D6> → Ir a Insertar > Insertar gráfico Proyección solar > Ajustar el título del gráfico: "Uso tradicional de plantas medicinales en Nacajuca, Tabasco" > Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico > Leyenda > derecha > → Etiqueta de datos > Centro> Más opciones de las etiquetas de datos> nombre de la categoría> valor> separador: nueva línea

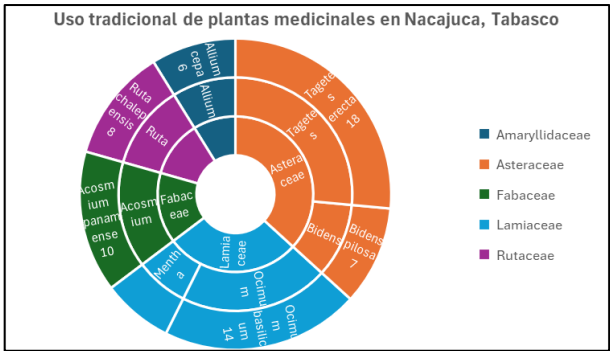


Figura. 4.23. Grafica de proyección solar

El gráfico muestra la jerarquía del centro hacia afuera, las familias, géneros y especies de plantas medicinales más usadas de Nacajuca, Tabasco, siendo el tamaño de cada arco proporcional al número de menciones de uso (Figura. 4.23.).

4.2.14. Gráfica de Cajas y bigotes

Como ingeniero ambiental, analizarás datos de altura de tres especies de mangle en Tabasco mediante un gráfico de caja y bigotes en Excel, identificando la mediana, la variabilidad y posibles valores atípicos. Este ejercicio te permitirá interpretar patrones dentro de rangos reales reportados para la región y aplicarlos en evaluaciones de ecosistemas costeros. (Cuadro.)

A1	:	x	✓	fx	Especie
	A			B	
1	Especie			Altura (m)	
2	Avicennia germinans L.			5	
3	Avicennia germinans L.			6.8	
4	Avicennia germinans L.			6.1	
5	Avicennia germinans L.			5.6	
6	Laguncularia racemosa (L.) Gaertn.			7.2	
7	Laguncularia racemosa (L.) Gaertn.			9.4	
8	Laguncularia racemosa (L.) Gaertn.			8.3	
9	Laguncularia racemosa (L.) Gaertn.			7.8	
10	Rhizophora mangle L.			8.9	
11	Rhizophora mangle L.			12	
12	Rhizophora mangle L.			10.1	
13	Rhizophora mangle L.			9.4	
14	Fuente: Corella Justavino, F. et. al 2001				

Cuadro

Abrir Excel >
Encabezados: A1:
Especie ; B1: Altura
(m) > Pegar datos
debajo> seleccionar
rango con
encabezados (A1:B1)>
Insertar > Gráficos
estadísticos > Caja y
bigotes> Título del
gráfico: "Distribución
de altura de mangles
en Tabasco" >

Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X: Especie); Horizontal secundario (Eje Y: Altura (m))>

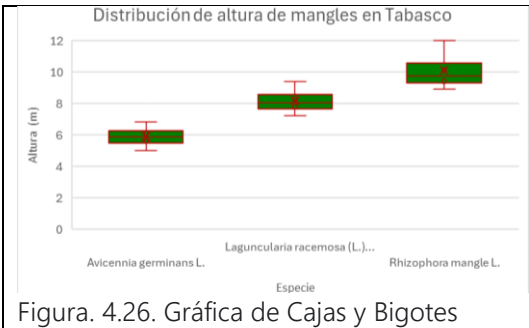


Figura. 4.26. Gráfica de Cajas y Bigotes

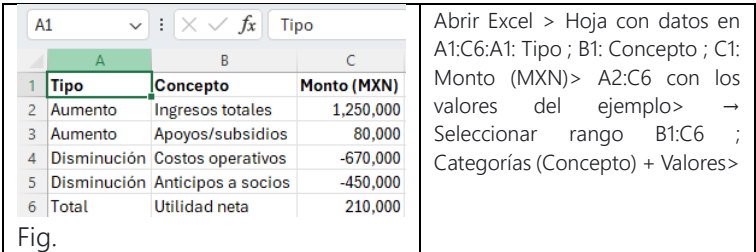


Figura. 4.27.

La X roja es la media, la línea en la caja es la mediana, la caja el rango típico y los bigotes la variación; aquí *Rhizophora* es la más alta, luego *Laguncularia* y *Avicennia* (Figura. 4.26.).

4.2.15. Gráfica de Cascada

Como gestor ambiental, usa una Gráfica de cascada para mostrar cómo los ingresos de una Cooperativa de Centla Tabasco (ventas, apoyos, turismo) se convierten en utilidad neta tras restar costos clave (combustible/hielo, mantenimiento, artes, permisos, anticipos). Si los valores son negativos representa dinero que sale, si el valor queda en positivo es dinero que queda. Así identificas impactos mayores y priorizas eficiencia y reducción de emisiones (Fig.).



Insertar > Gráficos > Cascada > Cascada>Título: Seleccionar el título > escribir: "Gráfica de cascada: ganancias y gastos de una cooperativa de Centla, Tabasco"> → En el gráfico, clic una vez en la columna "Utilidad neta" > clic derecho > Establecer como total> → Agregar etiquetas: Seleccionar gráfico > botón "+" (Elementos de gráfico) > Etiquetas de datos > Extremo externo>Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X: Concepto); Horizontal secundario (Eje Y: Monto (MXN))

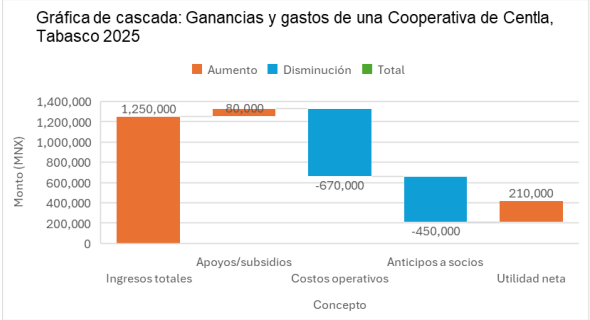


Figura. 4.28. Gráfica de Cascada

De izquierda a derecha: ingresos y apoyos suman, costos restan, y la barra final muestra la utilidad neta (210,000 MXN). Identificar la barra que baja más y trabajarla.

4.2.17. Gráfica de Combinado, ejemplo Climograma

Un gestor de Centro (Villahermosa) necesita ver juntos los patrones mensuales de lluvia (mm) y temperatura media (°C) para programar drenajes, riego y mantenimiento. Para resolverlo requiere un Climograma, para obtenerlo usa un gráfico combinado: columnas para lluvia y línea para temperatura en eje secundario; así identifica meses críticos (calor alto vs. lluvias intensas) y decide prioridades.

A1

:

f_x

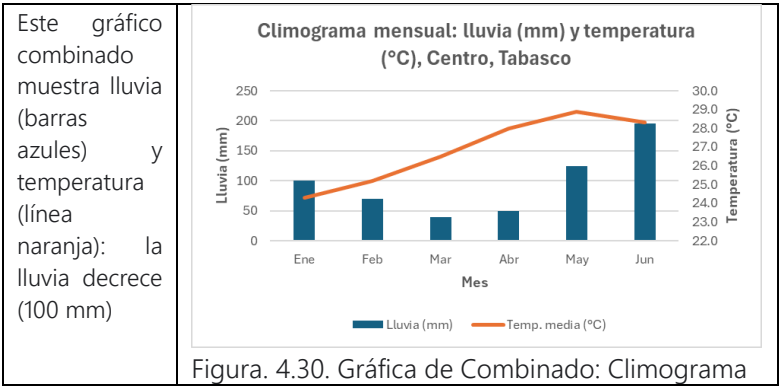
Mes

	A	B	C
1	Mes	Lluvia (mm)	Temp. media (°C)
2	Ene	100	24.3
3	Feb	70	25.2
4	Mar	40	26.5
5	Abr	50	28.0
6	May	125	28.9
7	Jun	195	28.3
8	Fuente: SMN-CONAGUA, 2025		

Cuadro.

Abrir Excel > Encabezados: A1: Mes; B1: Lluvia (mm); C1: Temp. media (°C)> A2:C6 con los valores del ejemplo> Seleccionar rango A1:C7> →Insertar > Gráficos > Combinado> Columna agrupada: línea en eje secundario>

Título del gráfico: escribir: "Climograma mensual: lluvia (mm) y temperatura (°C), Centro, Tabasco> Dale Clic dentro de la Gráfica> Diseño de gráfico (arriba en el Menú)> Agregar elemento de gráfico> Títulos del eje> Horizontal primario (Eje X: Mes); Vertical primario (Eje Y izquierdo: Lluvia (mm)); Vertical secundario (Eje Y derecho: Temperatura (°C))



4.2.17. Gráfica de Combinado, ejemplo Diagrama de Balance Hídrico, para detectar periodos de sequia

En Huimanguillo, la evapotranspiración supera la lluvia de en algunos meses, generando déficit hídrico y señalando sequía. ¿En qué mes el déficit es mayor y cómo afecta al ambiente

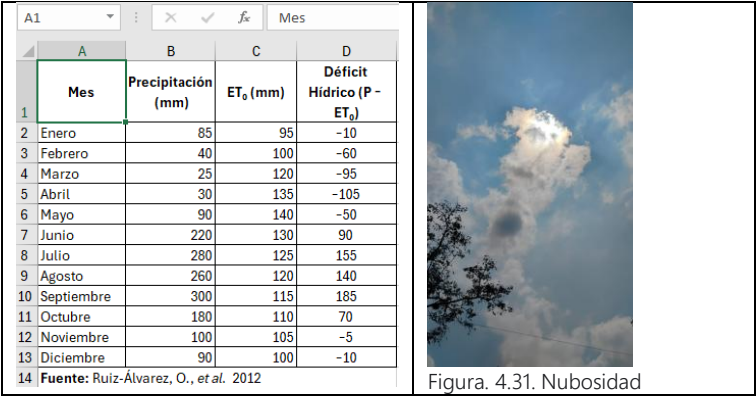


Figura. 4.31. Nubosidad

Ve a Insertar: Gráfico combinado, elige: Columna para Precipitación: Línea para ET₀. Línea o Área para Déficit Hídrico. Configura los ejes- Usa el eje primario para Precipitación y ET₀. Usa el eje secundario para el Déficit Hídrico si los valores están en otra escala.- Título: “Balance hídrico mensual – Huimanguillo, Tabasco”.

Las barras azules indican la lluvia, la línea naranja la evapotranspiración, y la verde el déficit hídrico. Cuando la evapotranspiración supera la precipitación, se identifica sequía; cuando ocurre lo contrario, hay disponibilidad de agua o bien si hay exceso de agua (Figura. 4.32.).

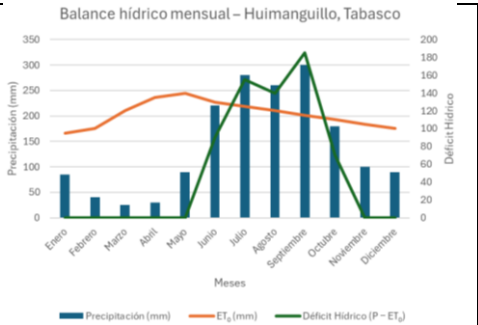


Figura. 4.32. Gráfica de Combinado: balance hídrico mensual

5. Estadística inferencial

Estimación de parámetros

La estimación de parámetros es el proceso de inferir un valor desconocido (parámetro) de una población utilizando datos de una muestra (ver 4.1).

Intervalos de confianza

47

Un intervalo de confianza es una herramienta estadística que sirve para dar un rango de valores dentro del cual es muy probable que esté el verdadero parámetro de la población (como la media o la proporción).

En lugar de dar solo una estimación puntual (ejemplo: la media muestral), se construye un intervalo que refleja la incertidumbre del cálculo.

Incertidumbre

Por qué antes de probar hipótesis, hay que saber cómo estimar parámetros y medir la incertidumbre, para ello utilizamos:

Error estándar: Es una medida que nos dice qué tanto puede variar el resultado de una muestra respecto al verdadero valor de la población. Mientras más grande sea la muestra, más pequeño suele ser este error.

Desviación estándar / varianza: Nos muestran qué tan dispersos están los datos. Si los datos están muy separados entre sí, hay más incertidumbre; si están más juntos, hay menos.

Intervalos de confianza: En lugar de dar un solo número como estimación, damos un rango. Ese rango refleja la incertidumbre y nos dice: "con un 95% de seguridad, el valor real está dentro de este intervalo".

Pruebas de hipótesis y nivel de significancia

Pruebas de hipótesis: Son procedimientos estadísticos que usamos para decidir si una idea (hipótesis) sobre una población es razonable o no, basándonos en datos de una muestra. Por ejemplo: "¿los estudiantes de una escuela tienen en promedio más de 20 años?".

Nivel de significancia: Es el límite que ponemos antes de analizar los datos para decidir si rechazamos la hipótesis nula. Normalmente se usa 5% (0.05), lo que significa aceptar un pequeño riesgo de equivocarnos al decir que hay diferencia cuando no existe (de 100 pruebas, 5 podrían ser errores).

La Hipótesis nula (H_0) y alternativa (H_a)

Hipótesis nula (H_0): Es como el "punto de partida" que asumimos hasta que los datos digan lo contrario. En estudios ambientales suele plantear que no hay cambio o no hay diferencia. Ejemplo: "La calidad del agua en el río es igual antes y después de la descarga de la planta".

Hipótesis alternativa (H_a): Es la idea que queremos comprobar, la que plantea que sí hay un efecto o sí hay diferencia. Ejemplo: "La calidad del agua en el río cambió después de la descarga de la planta".

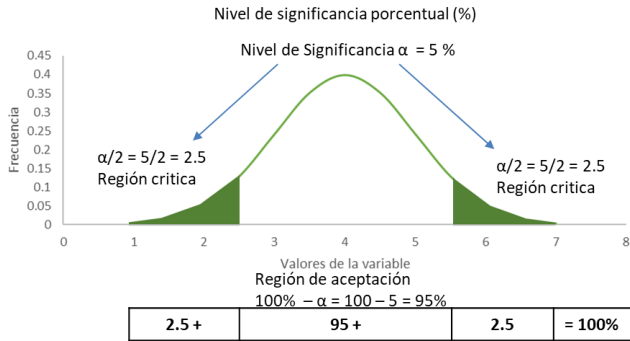


Figura 5.1. Nivel de significancia porcentual

Hipótesis Nula (H_0): ($A = B$)

Indica que No hay diferencia significativa entre A y B; se asume que son iguales.

Hipótesis Alternativa (H_a): ($A \neq B$)

Indica que hay una diferencia significativa entre A y B; se asume que no son iguales.

Cálculo del Valor Z:

Se calcula el valor Z (valor crítico) usando los datos obtenidos en campo y se compara con el estadístico de prueba.

Regla de Decisión:

Si los resultados caen dentro del área del 95%, no rechazamos (H_0).

Si caen en las áreas críticas, rechazamos (H_0) y aceptamos (H_a).

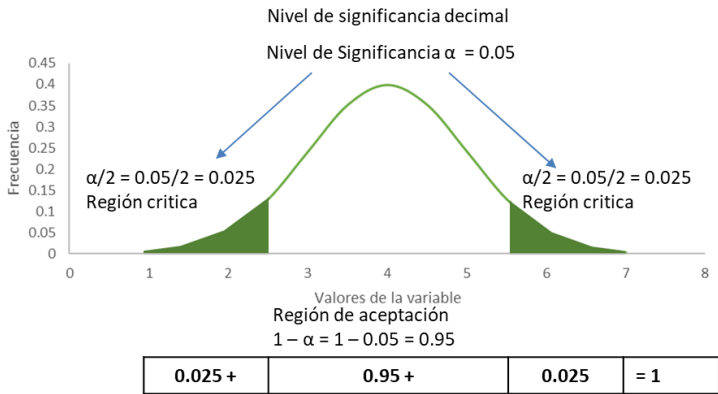


Figura 5.2. Nivel de significancia decimal

Los dos gráficos representan el mismo concepto, pero de diferentes maneras (porcentual y decimal). El uso de decimales es común en cálculos estadísticos, mientras que los porcentajes son más fáciles para la interpretación y redacción de resultados para el público que usara los resultados.

El valor p, su relación con α y error tipo I

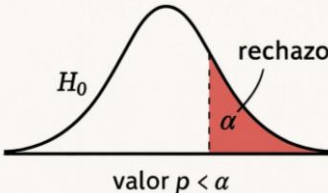
El Valor p: Es un número que nos dice qué tan fuertes son las evidencias contra la hipótesis nula. Mientras más pequeño sea el valor p, más evidencia tenemos de que la hipótesis nula probablemente no es cierta.

Relación entre p y α : Si el valor p es menor que α , decimos que los datos son suficientemente fuertes para rechazar la hipótesis nula. Si el valor p es mayor, no la rechazamos.

Error tipo I: Es el riesgo de equivocarnos al rechazar la hipótesis nula cuando en realidad era cierta. Ese riesgo es justamente el nivel de significancia α

Error tipo II (β): Es cuando no detectamos un efecto que sí existe. Es decir, la hipótesis nula en realidad es falsa, pero los datos no nos dieron suficiente evidencia para rechazarla. Ejemplo ambiental: el agua sí está contaminada, pero el análisis concluye que "no hay diferencia".

Poder estadístico ($1 - \beta$): Es la capacidad de una prueba para detectar un efecto real cuando existe. Mientras mayor sea el poder, más probable es que descubramos cambios o diferencias verdaderas.

Error de tipo I y valor p  Figura 5.3. Error tipo I y valor p	Error tipo I y valor p La curva representa la hipótesis nula (H_0), es decir, que no hay diferencia. La zona roja al final de la curva es la zona de rechazo: si los datos caen ahí, rechazamos H_0. El valor p es la probabilidad de que los datos caigan en esa zona. Si $p < \alpha$ (por ejemplo, menor a 0.05), se rechaza H_0. Pero al hacerlo, existe el riesgo de cometer un error tipo I: decir que hay diferencia cuando en realidad no la hay (Figura 5.3.).
Error de tipo II y poder estadístico  Figura 5.4. Error tipo II	Error tipo II y poder estadístico Hay dos curvas: una para la hipótesis nula (H_0) y otra para la alternativa (H_a), que dice que sí hay diferencia (Figura. 5.4.).

	La zona verde (β) es el error tipo II: no detectar una diferencia que sí existe. La zona azul es el poder estadístico: la capacidad de detectar correctamente esa diferencia.
--	--

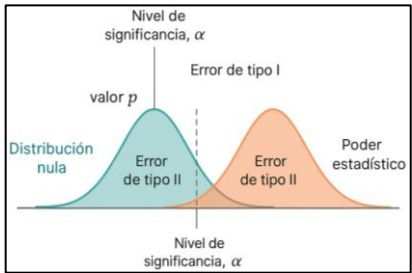


Figura 5.5. Error tipo I y II

La imagen muestra la prueba de hipótesis y los conceptos de error tipo I y error tipo II:

Curvas de Distribución: Hay dos curvas (Figura 5.5.)

Distribución Nula (azul-verde): Representa la hipótesis nula (H_0).

Distribución Alternativa (naranja): Representa la hipótesis alternativa (H_a).

Error Tipo I: Ocurre en la cola derecha de la distribución nula, donde se rechaza (H_0) cuando es verdadera (falso positivo).

Error Tipo II: Se encuentra en la cola izquierda de la distribución alternativa, donde no se rechaza (H_0) cuando (H_a) es verdadera (falso negativo).

Nivel de Significancia (α): Línea vertical que indica el umbral para decidir si se rechaza (H_0).

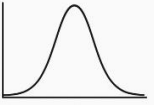
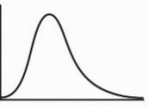
Valor p: Probabilidad de observar los resultados, ayuda a tomar decisiones sobre (H_0).

Poder Estadístico: Capacidad de detectar un efecto real, relacionado con la probabilidad de rechazar correctamente (H_0).

5.1. Estadística paramétrica

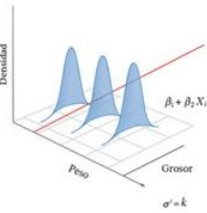
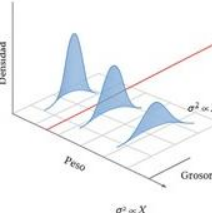
5.1.1. Normalidad

La normalidad en estadística indica que los datos siguen una distribución simétrica con forma de campana, llamada distribución normal o gaussiana.

		
Figura. 5.6. Asimetría izquierda	Figura. 5.7 Normalidad	Figura. 5.8. Asimetría derecha

5.1.2. Homocedasticidad

La homocedasticidad significa que la variabilidad de los datos es constante entre grupos o condiciones, permitiendo comparaciones justas en análisis estadísticos.

Homocedasticidad  Figura. 5.9. Las varianzas son similares: Ideal para comparar grupos con estadística Paramétrica	Heterocedasticidad  Figura. 5.10. Las varianzas son diferentes: Su comparación es con estadística no Paramétrica
---	---

Homocedasticidad: en Densidad y Peso

1. Variación constante del peso: Los peces de la misma talla tienen pesos que varían igual en todo el rango.
2. Dispersión uniforme de la densidad: Las células mantienen la misma variabilidad de densidad sin importar el tamaño del tejido.
3. Error consistente en las mediciones: Al medir peso en ratones, el margen de error es igual para ratones de 20g que para ratones de 40g.

5.1.3. Selección de Técnicas Estadísticas para analizar datos

Después de cumplir los supuestos de Normalidad y Homocedasticidad, según sea el caso, elige la técnica a utilizar para tus análisis de datos. Identifica que es lo que quieres. Mira tus datos y completa esta frase:

"Quiero CONOCER mis datos" → Describir

"Quiero COMPARAR grupos" → Comparar

"Quiero CONECTAR dos variables" → Relacionar

"Quiero CALCULAR el futuro" → Predecir

"Quiero CREAR algo especial" → Especiales

55

Con cuál de ellas identificas tu actividad, si aun tienes dudas de cuál es, mira estos ejemplos:

DESCRIBIR = usa estadística descriptiva (Ver el punto 4)

¿Cuál es el promedio de contaminación en mi laguna?

¿Cómo se distribuyen los pesos de las tortugas?

COMPARAR = ¿Hay diferencias entre grupos?

¿Los peces de agua dulce son más grandes que los de agua salada?

¿La zona protegida tiene más biodiversidad?

RELACIONAR = Van juntas dos cosas

¿A más temperatura, menos oxígeno disuelto?

¿El tamaño del árbol se relaciona con su edad?

PREDECIR = ¿Qué pasará después?

¿Cuánta lluvia habrá el próximo mes?

¿Cómo crecerá la población de aves?

ESPECIALES = Necesito algo único

Tomar muestra representativa de peces

Elegir 20 parcelas al azar de 200

5.1.4. Técnicas estadísticas para Describir

Atributos del grupo: no requiere normalidad y está pensado para datos numéricos (mediciones).

1. ¿Tus datos son numéricos (medidas, conteos)?

Sí → 2.

No → Este grupo no aplica (son nombres, frecuencias, usa tablas/gráficos de categorías). 4.1.

2. ¿Quieres un resumen numérico (media, mediana, desviación, asimetría, curtosis, n, IC)?

Sí → Estadística descriptiva. 4.1.

No → 3.

3. ¿Quieres ver la forma de la distribución (montículos, colas, valores extremos, datos atípicos) en clases? Puede ser la longitud de hojas de cacao (*Teobroma cacao* L.)

Sí → Histograma.

No → 4.

4. ¿Quieres ordenar y posicionar observaciones por su lugar (clasificación, percentiles, deciles, cuartiles)? El ruido medido en Decibeles (DB) en un entorno laboral, por ejemplo

Sí → Jerarquía y percentil.

No → Replantea objetivo (quizá Comparar o Relacionar).

5.1.4.1. Estadística descriptiva

Ver el punto 4.1. y el 4.2.

Nota: Los datos utilizados en los ejemplos de este capítulo son simulados y se emplean exclusivamente con fines didácticos y de aprendizaje estadístico.

5.1.4.2. Histograma

El histograma es una gráfica que muestra cómo se distribuyen los datos, agrupándolos en intervalos y representando con barras la frecuencia de cada intervalo.

Uso: analizar patrones y variabilidad.

Ejemplo: medir longitudes de hojas (cm) de Cacao, ver en qué rangos se concentran la mayoría.

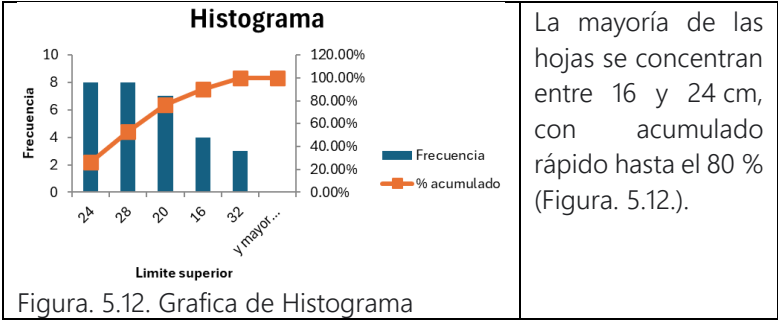
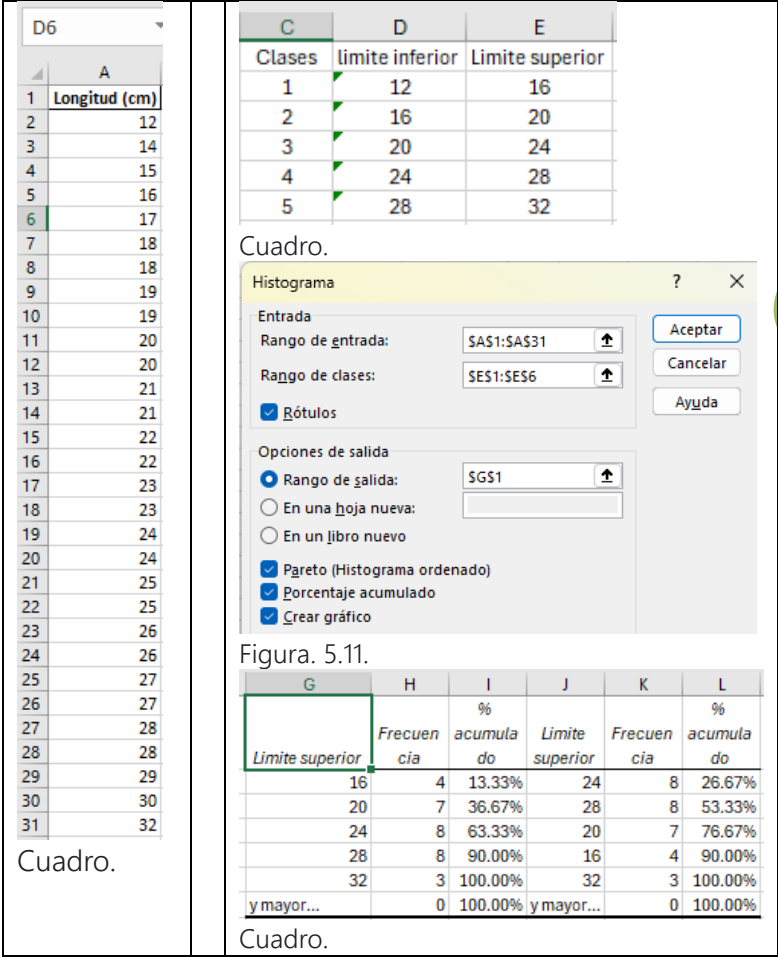
1. Activa el complemento: Archivo: Opciones: Complementos: Ir; Activa Herramientas para análisis. En la pestaña Datos aparecerá Análisis de datos.

2. Prepara tus columnas: Una columna con tus datos (ejemplo: longitud cm; Para calcular el número de clases, usa la raíz cuadrada del número de datos (30), ejemplo: raíz de 30, es 5.47, redondeado serían 5 clases: la diferencia del mayor (32) y menor (12) valor es 20, se divide entre 5 y da 4: Otra columna con los intervalos (ejemplo: 12–16, 16–20).

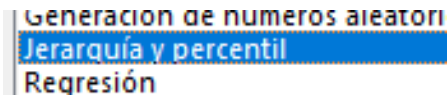
3. Abre Histograma: En Datos: Análisis de datos, selecciona Histograma, da Aceptar.

4. Configura entradas: Rango de entrada: tus datos; Rango de intervalos". Marca Etiquetas si tu primera fila tiene títulos; Elige un lugar sin celdas ocupadas, para la salida de datos y el gráfico.

5. Genera el resultado: Obtendrás una tabla con frecuencias; Si marcas las tres últimas opciones, con el Gráfico de salida, verás el histograma automáticamente.



5.1.4.3. Jerarquía y percentil

En una planta de agua en Villahermosa, Tabasco, se han medido los niveles de ruido de las bombas a lo largo del día. El ruido puede variar en diferentes horas, y es importante monitorear estos niveles para cumplir con la NOM 011 STPS 2001, que exige protección auditiva al superar los 85 dB(A).																																						
1. Activa el complemento (no es necesario si ya fue activado): Archivo; Opciones; Complementos; Ir; Activa "Herramientas para análisis"	<table><tr><th></th><th>A</th><th>B</th></tr><tr><td>1</td><td>Hora</td><td>Nivel de Ruido dB(A)</td></tr><tr><td>2</td><td>08:00 a. m.</td><td>82</td></tr><tr><td>3</td><td>10:00 a. m.</td><td>88</td></tr><tr><td>4</td><td>12:00 p. m.</td><td>90</td></tr><tr><td>5</td><td>02:00 p. m.</td><td>85</td></tr><tr><td>6</td><td>04:00 p. m.</td><td>92</td></tr><tr><td>7</td><td>06:00 p. m.</td><td>78</td></tr><tr><td>8</td><td>08:00 p. m.</td><td>84</td></tr><tr><td>9</td><td>10:00 p. m.</td><td>91</td></tr><tr><td>10</td><td>12:00 a. m.</td><td>87</td></tr><tr><td>11</td><td>02:00 a. m.</td><td>89</td></tr></table> Cuadro.		A	B	1	Hora	Nivel de Ruido dB(A)	2	08:00 a. m.	82	3	10:00 a. m.	88	4	12:00 p. m.	90	5	02:00 p. m.	85	6	04:00 p. m.	92	7	06:00 p. m.	78	8	08:00 p. m.	84	9	10:00 p. m.	91	10	12:00 a. m.	87	11	02:00 a. m.	89	
	A	B																																				
1	Hora	Nivel de Ruido dB(A)																																				
2	08:00 a. m.	82																																				
3	10:00 a. m.	88																																				
4	12:00 p. m.	90																																				
5	02:00 p. m.	85																																				
6	04:00 p. m.	92																																				
7	06:00 p. m.	78																																				
8	08:00 p. m.	84																																				
9	10:00 p. m.	91																																				
10	12:00 a. m.	87																																				
11	02:00 a. m.	89																																				
2. Ingresas los datos: En una hoja nueva, escribe la tabla: primera columna Hora; segunda columna Nivel de Ruido dB(A)																																						
3. Abre Análisis de datos; Pestaña Datos; Análisis de datos: elige "Jerarquía y percentil".	 Figura. 5.13. Opción Jerarquía y percentil																																					

4. Configura; Rango de entrada: columna de ruido (Nivel de Ruido dB(A)); Rango de salida: celda donde quieres los resultados. Marca “Rótulos” si incluiste encabezado.

Jerarquía y percentil

Entrada

Rango de entrada:

Agrupado por: ☒ Columnas ☐ Filas

☒ Rótulos en la primera fila

Opciones de salida

☒ Rango de salida:

☐ En una hoja nueva:

☐ En un libro nuevo

Aceptar

Cancelar

Ayuda

Figura. 5.14. Cuadro de diálogo, Jerarquía y percentil

5. Resultado; Excel genera una tabla con los datos ordenados y su percentil.

	A	B	C	D	E	F
	Hora	Nivel de Ruido dB(A)	Posición	Nivel de Ruido dB(A)	Jerarquía	Porcentaje
1						
2	08:00 a. m.	82	5	92	1	100.00%
3	10:00 a. m.	88	8	91	2	88.80%
4	12:00 p. m.	90	3	90	3	77.70%
5	02:00 p. m.	85	10	89	4	66.60%
6	04:00 p. m.	92	2	88	5	55.50%
7	06:00 p. m.	78	9	87	6	44.40%
8	08:00 p. m.	84	4	85	7	33.30%
9	10:00 p. m.	91	7	84	8	22.20%
10	12:00 a. m.	87	1	82	9	11.10%
11	02:00 a. m.	89	6	78	10	0.00%

Cuadro.

En la planta de agua en Villahermosa se midieron niveles de ruido entre 78 y 92 dB(A). El valor más alto fue a las 04:00 PM, representa el 55.50% del total de ruido, lo que indica horas críticas que requieren protección auditiva. Con jerarquía se ordenan los niveles y con percentiles se analiza la distribución.

Puedes elaborar un gráfico de columnas e incluir una línea donde se muestre el valor de 80 dB(A) que es el límite que

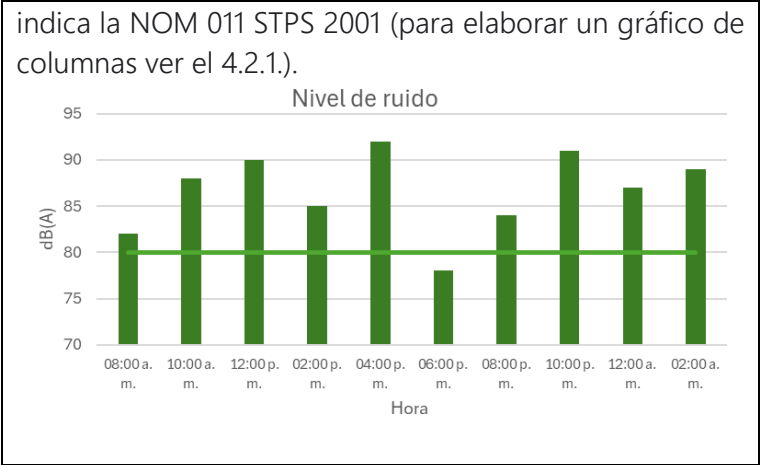


Figura: 5.15. Gráfico de columnas, para ver Jerarquía y percentil

5.1.5. Técnicas Estadísticas Paramétricas para comparar grupos

1. ¿Quieres comparar varianzas de dos grupos independientes (sujetos diferentes)?

Sí → Prueba F para varianzas de dos muestras.

No → 2.

2. ¿Cuántos grupos compararás en la media?

Dos → 3.

Más de dos ($k \geq 3$) → Análisis de varianza de un factor.

3. Comparas dos grupos de datos, ¿Las mediciones son pareadas (antes–después con las mismas unidades)?

Sí → Prueba t para medias de dos muestras emparejadas.

No, son independientes → 4.

4. ¿Conoces la desviación estándar de la población en ambos grupos (σ_1 y σ_2)?

Sí → Prueba z para medias de dos muestras.

No → 5.

5. Dos grupos independientes y la desviación estándar poblacional (σ) desconocida: ¿las varianzas son similares?

Paso 1, comprobar igualdad de varianzas con la Prueba F para varianzas de dos muestras, si el valor de p es > 0.05

Sí → Prueba t para dos muestras suponiendo varianzas iguales.

No → el valor de p es ($p \leq 0.05$), Prueba t para dos muestras suponiendo varianzas desiguales.

6. Si hay dos factores (p.ej., sitio $\times$ mes): ¿tienes replicación (≥ 2 por celda)?

Sí → Análisis de varianza de dos factores con varias muestras por grupo.

No → Análisis de varianza de dos factores con una sola muestra por grupo.

Emparejadas (o Pareadas): mismas unidades medidas antes–después o parejas 1:1 comparables

En la siguiente sección se desarrollarán ejemplos de estadística inferencial, sin embargo, es importante aclarar que.

En este documento se emplean artículos científicos únicamente como referencia contextual para ilustrar ejercicios de estadística inferencial. Los datos presentados en los ejemplos son ficticios y creados con fines educativos, sin relación directa con los resultados originales de los artículos citados. Cualquier coincidencia con información real es accidental. Este enfoque permite comprender y aplicar los conceptos estadísticos sin copiar datos específicos de investigaciones previas.

5.1.5.1. Prueba F para varianzas de dos muestras

En el contexto de la degradación de los recursos naturales, es fundamental estudiar la variabilidad en la cobertura vegetal entre áreas con diferentes tipos de manejo en el Parque Estatal Agua Blanca, Tabasco. Este análisis es crucial para comprender los efectos de la intervención humana en la biodiversidad (Castillo Acosta & Zavala Cruz, 1996). A continuación, se presenta un ejercicio hipotético en el que se investiga la variabilidad de la cobertura vegetal entre áreas de conservación y explotación. Se utilizará la Prueba F para varianzas de dos muestras, adecuada para comparar las varianzas de dos grupos independientes y determinar diferencias significativas. La pregunta de investigación es: ¿Existen diferencias significativas en la varianza de la cobertura vegetal entre áreas de conservación y explotación en el Parque Estatal Agua Blanca?

B. Datos: datos de cobertura vegetal (Cuadro).

A		B	
1	Cobertura Vegetal (m²), en Conservación	2	Cobertura Vegetal (m²), en Explotación
2	45	3	30
3	50	4	32
4	48	5	28
5	52	6	35
6	47	7	31

Cuadro.

C. Plantear Hipótesis

Hipótesis Nula (H_0) No hay diferencias significativas en la varianza de la cobertura vegetal entre áreas de conservación y explotación. $H_{01}: \sigma^2_i = \sigma^2_j$

Hipótesis Alternativa (H_a): Hay diferencias significativas en la varianza de la cobertura vegetal entre áreas de conservación y explotación. $H_{a1}: \sigma^2_i \neq \sigma^2_j$.

D. Regla de Decisión: Si el valor p es menor (<) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

Menor (<) ; Mayor (>)

E. Establecer el Nivel de Significancia (α : alfa):

Se utilizará un nivel de significancia del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Prueba F para varianzas de dos muestras

Prueba F para varianzas de dos muestras

Entrada

Rango para la variable 1:

Rango para la variable 2:

☒ Rótulos

Alfa:

Opciones de salida

☒ Rango de salida:

☐ En una hoja nueva:

☐ En un libro nuevo

Aceptar

Cancelar

Ayuda

Primer columna

Segunda columna

Rótulos

Alfa (α) de 0.05

Salida de datos

Aceptar

Figura. 5.16. Cuadro de diálogo, Prueba F para varianzas de dos muestras

G. Resultados del Analisis de la Prueba F para varianzas de dos muestras

Prueba F para varianzas de dos muestras		
	Cobertura Vegetal (m ²), en Conservación	Cobertura Vegetal (m ²), en Explotación
Media	48.4	31.2
Varianza	7.3	6.7
Observaciones	5	5
Grados de libertad	4	4
F	1.089552239	
P(F<=f) una cola	0.467876822	
Valor critico para F (una cola)	6.388232909	

Cuadro.

H. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor (<) que (α) 0.05, se rechazará la (H_0).

H1 (resultado en filas): Valor p = 0.46 > (α) 0.05, por lo tanto, se acepta la (H_a).

I. Resultado del análisis: Los resultados muestran un p-valor de 0.467876, indicando que no hay diferencias significativas en la varianza de la cobertura vegetal entre conservación y explotación. Esto sugiere que se debe reforzar la conservación.

5.1.5.2. Análisis de varianza de un factor.

A. Contexto: En el sureste de México, la producción de cacao (*Theobroma cacao* L.) enfrenta pérdidas superiores al 75% por moniliasis (*Moniliophthora roreri*) (Figura. 5.17.), Torres-De la Cruz *et al.* (2020) reportan una incidencia promedio de 75.8% en la región. El estudio se desarrolla en tres plantaciones (Tapijulapa, Oxolotán y Zunu y Patastal) durante el ciclo 2023-2024, con seis muestreos por plantación. La pregunta de investigación es: ¿Existen diferencias significativas en la incidencia de moniliasis entre las plantaciones de cacao? Este ejemplo corresponde a un ANOVA de un factor, donde el factor es la plantación y la variable de respuesta la incidencia de la enfermedad.

67

B. Datos: datos sobre la incidencia de moniliasis (Cuadro).

	A	B	C	
	Incidencia Tapijulapa (%)	Incidencia Oxolotan (%)	Incidencia Zunu y Patastal (%)	
1				
2	82	70	75	
3	78	72	78	
4	80	68	73	
5	75	71	76	
6	81	69	77	
7	79	74	74	



Cuadro.

Figura. 5.17.

C. Plantear Hipótesis

Hipótesis Nula (H_0): No hay diferencias significativas en la incidencia de moniliasis según la plantación (las medias son iguales: $\mu_i = \mu_j$).

Hipótesis Alternativa (H_a): Hay diferencias significativas en la incidencia de moniliasis según la plantación (al menos una media es diferente: $\mu_i \neq \mu_j$).

D. Regla de Decisión: Si el valor p es menor (<) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

Menor (<) ; Mayor (>)

E. Establecer el Nivel de Significancia (α : alfa):

Se utilizará un nivel de significancia (α) del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Análisis de varianza de un factor

Analisis de varianza de un factor

Entrada

Rango de entrada

SAS1:SC57

Aceptar

Cancelar

Agrupado por:

☒ Columnas

☐ Filas

Ayuda

☒ Rótulos en la primera fila

Alfa:

0.05

Opciones de salida

☒ Rango de salida:

SAS9

☐ En una hoja nueva:

☐ En un libro nuevo

Selecciona todo

En columnas

Rótulos (✓), Alfa o valor de significancia (α) de 0.05

Salida de datos

Dale Aceptar

Figura. 5.18.

G. Resultados del Análisis de Varianza (ANOVA) de un Factor)

	A	B	C	D	E	F	G
11	RESUMEN						
12	Grupos	Cuenta	Suma	Promedio	Varianza		
13	Incidencia Tapijulapa (%)	6	475	79.16666667	6.16666667		
14	Incidencia Oxolotan (%)	6	424	70.66666667	4.66666667		
15	Incidencia Zunu y Pastatal (%)	6	453	75.5	3.5		
16				D19/D20			
17	ANÁLISIS DE VARIANZA		F (2, 15)	109.05/4.7~22.83			
18	Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor critico para F
19	Entre grupos	218.1111111	2	109.0555556	22.8255814	2.81443E-05	3.682320344
20	Dentro de los grupos	71.66666667	15	4.777777778			
21							
22	Total	289.7777778	17				

Cuadro.

H. Contraste de Hipótesis: Según la regla de decisión: Si p calculado es menor(<) que (α) 0.05, se rechazará la Ho.

Valor $p = 2.81 \times 10^{-5} < 0.05$, por lo tanto, se rechaza la Ho

I. Resultado del análisis: Sí, existen diferencias significativas en la incidencia de moniliasis entre las plantaciones de cacao evaluadas.

Con un nivel de significancia de $\alpha=0.05$, el ANOVA de un factor mostró diferencias significativas en la incidencia de moniliasis entre las tres localidades evaluadas (Tapijulapa, Oxolotán y Zunu-Pastatal), $F(2,15)=22.83$, $p=2.81 \times 10^{-5} < 0.05$.

5.1.5.3. Prueba t para medias de dos muestras emparejadas

Esta prueba es también conocida como t de Student para muestras relacionadas.

En la Laguna de las Ilusiones, Villahermosa, Tabasco, se ha observado un aumento en el diámetro basal de los nidos de cocodrilo de pantano (*Crocodylus moreletii*), lo que podría reflejar cambios significativos en el hábitat urbano (López-Luna *et al.* 2011). El siguiente ejercicio utiliza un escenario supuesto para analizar cómo una intervención en el manejo de la orilla influye en dicho diámetro. Durante mayo-julio de 2026, un equipo de Biólogos evaluará 10 sitios. La pregunta de investigación es: ¿Existe un cambio significativo en el diámetro de los nidos después de la intervención en la orilla? Se utilizará la prueba t pareada para evaluar las mediciones del mismo sitio.

B. Datos: datos de diámetro de nidos (Cuadro).

	A	B	C	
1	Sitio	Antes (cm)	Después (cm)	
2	1	152	162	
3	2	160	167	
4	3	147	159	
5	4	172	181	
6	5	165	173	
7	6	170	181	
8	7	158	164	
9	8	175	188	
10	9	162	171	
11	10	169	179	



Figura 5.19. Nido de cocodrilo

Cuadro.

C. Plantear Hipótesis

Hipótesis Nula (H_0): No existe diferencia en el diámetro basal medio de los nidos antes y después de la intervención. $H_0: \mu_i = \mu_j$

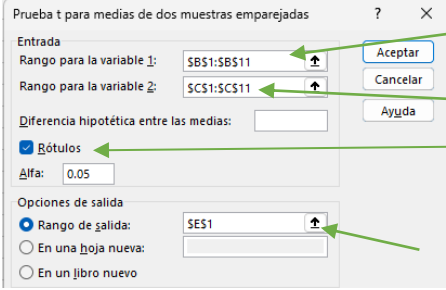
Hipótesis Alternativa (H_a): Existe una diferencia en el diámetro basal medio de los nidos antes y después de la intervención: $\mu_i \neq \mu_j$).

D. Regla de Decisión: Si el valor p es menor (<) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

Menor (<) ; Mayor (>)

E. Establecer el Nivel de Significancia (α : alfa):
Se utilizará un nivel de significancia (α) del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Prueba t para medias de dos muestras emparejadas



Selecciona el antes

Selecciona el después

Rótulos (✓), Alfa (α) o valor de significancia de 0.05

Salida de datos
Dale Aceptar

Figura. 5.20. Cuadro de diálogo, Prueba t para medias de dos muestras emparejadas

G. Resultados del Análisis de Prueba t para medias de dos muestras emparejadas

E	F	G
Prueba t para medias de dos muestras emparejadas		
	Antes (cm)	Después (cm)
Media	163	172.5
Varianza	80.6666667	91.61111111
Observaciones	10	10
Coefficiente de correlación de Pearson	0.97455803	
Diferencia hipotética de las medias	0	
Grados de libertad	9	
Estadístico t	-13.824531	
P(T<=t) una cola	1.1434E-07	
Valor crítico de t (una cola)	1.83311293	
P(T<=t) dos colas	2.2868E-07	
Valor crítico de t (dos colas)	2.26215716	

$2.28 \times 10^{-7} = 0.00000022868$

$2.2868E-07 = 2.28 \times 10^{-7}$

H. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor ($<$) que (α) 0.05, se rechazará la (H_0).
Valor $p = 2.28 \times 10^{-7} < 0.05$, por lo tanto, se rechaza la (H_0).

I. Resultado del análisis: El análisis reveló un estadístico t de -13.82 y un p-valor bilateral de 2.28×10^{-7} , menor que el valor de significancia de 0.05, confirma que existe un cambio significativo en el diámetro de los nidos de cocodrilo tras la intervención en la orilla.

5.1.5.4. Prueba z para medias de dos muestras.

Se necesitan 30 datos mínimos por grupo; en el ejemplo serán 15.

En Tabasco, México, ante la necesidad de seguridad alimentaria y los hongos representan una alternativa (Carreño-Ruiz et al., 2020). En un supuesto se evaluará la producción de *Schizophyllum commune* entre enero y marzo de 2026. Mediante una prueba Z para medias de dos muestras, se comparará la Eficiencia Biológica (EB %) obtenida en cáscara de Pachira frente a hojas de Cocohite bajo la siguiente interrogante: ¿Difiere significativamente la Eficiencia Biológica media en la producción de *Schizophyllum* al comparar el uso de cáscara de Pachira frente a hojas de Cocohite?

71

B. Datos: datos de Eficiencia Biológica (EB %) (Cuadro).

A		B	C
ID	Eficiencia Biológica (EB) - Cáscara de Pachira (%)	Eficiencia Biológica (EB) - Hojas de Cocohite (%)	
1			
2	1	12.5	8.4
3	2	13.1	9.0
4	3	11.8	7.5
5	4	12.0	8.2
6	5	13.5	9.3
7	6	12.8	8.7
8	7	11.9	7.9
9	8	12.3	8.5
10	9	13.0	9.1
11	10	12.7	8.3
12	11	13.2	9.4
13	12	11.6	7.7
14	13	12.1	8.0
15	14	13.4	9.2
16	15	12.4	8.6
17	Varianza	0.34	0.33
18	=VARP(B2:B16)		=VARP(C2:C16)



Figura. 5.21.
Schizophyllum

C. Plantear Hipótesis

Hipótesis Nula (H_0): No hay diferencia en la Eficiencia Biológica entre los sustratos de cáscara de Pachira y hojas de Cocohite. $H_0: \mu_i = \mu_j$

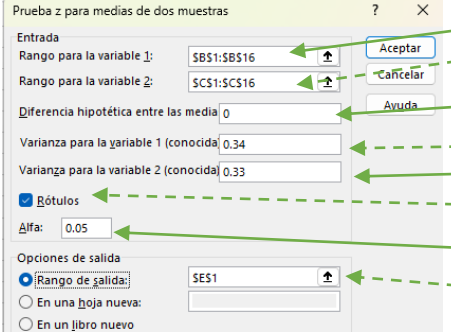
Hipótesis Alternativa (H_a): Hay una diferencia en la Eficiencia Biológica entre los sustratos de cáscara de Pachira y hojas de Cocohite. $H_a: \mu_i \neq \mu_j$.

D. Regla de Decisión: Si $p \text{ valor} < 0.05$, se rechaza H_0

también se puede usar, Si $|Z \text{ calculado}| > Z \text{ critico}$ se rechaza H_0
Menor ($<$) ; Mayor ($>$)

E. Establecer el Nivel de Significancia:
Se utilizará un nivel de significancia del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Prueba z para medias de dos muestras



Prueba z para medias de dos muestras

Entrada

Rango para la variable 1: $\$B\$5:\$B\16

Rango para la variable 2: $\$C\$5:\$C\16

Diferencia hipotética entre las media: 0

Varianza para la variable 1 (conocida): 0.34

Varianza para la variable 2 (conocida): 0.33

☒ Rótulos

Alfa: 0.05

Opciones de salida

☒ Rango de salida: $\$E\5

☐ En una hoja nueva:

☐ En un libro nuevo

Selecciona Pachira

Selecciona Cocohite

Diferencia, es cero (0)

Varianza de Pachira

Varianza de Cocohite

Rótulos (✓)

Alfa de 0.05

Salida de datos

Dale Aceptar

G. Resultados del Análisis de Prueba z para medias de dos muestras

E	F	G	H
Prueba z para medias de dos muestras			
	Eficiencia Biológica (EB) - Cáscara de Pachira (%)	Eficiencia Biológica (EB) - Hojas de Cocohite (%)	
Media	12.5533333	8.52	
Varianza (conocida)	0.34	0.33	
Observaciones	15	15	
Diferencia hipotética de las medias	0		
z	19.084129		Z calculado
P(Z<=z) una cola	0		
Valor critico de z (una cola)	1.64485363		
Valor critico de z (dos colas)	0		p-valor
Valor critico de z (dos colas)	1.95996398		Z critico

informacion agregada

H. Contraste de Hipótesis: Según la regla de decisión: Si el p valor es menor ($<$) que (α) 0.05, se rechazará la (H_0).

Valor $p = 0.0 < 0.05$, por lo tanto, se rechaza la (H_0).

La otra regla: $Z \text{ calculado} (19.08) > Z \text{ critico} (0)$, se rechaza H_0 .

I. Resultado del análisis: El análisis mostró un estadístico Z de 19.08 y un p-valor bilateral de 0.0, que es menor que el nivel de significancia (α) de 0.05. Esto confirma que existe una diferencia significativa en la Eficiencia Biológica entre el uso de cáscara de Pachira ($\bar{x}=12.5$) y las hojas de Cocohite; ($\bar{x}=8.52$).

5.1.5.5. Prueba t para dos muestras suponiendo varianzas iguales

El estudio de la contaminación del suelo por metales pesados es crucial para la seguridad alimentaria y la salud ambiental (Álvarez-Coronel *et al.* 2023). Este análisis se apoya en un escenario simulado para determinar los efectos de la contaminación en la productividad agrícola en el Activo Samaria-Luna, Cunduacán, Tabasco, durante 2026. La pregunta es ¿Existe un cambio significativo en el rendimiento de cultivos en suelos contaminados por metales pesados en comparación con suelos no contaminados? utilizando la Prueba t para dos muestras suponiendo varianzas iguales.

B. Datos: datos de rendimientos de maíz (kg/ha) (Cuadro).

A		B	
	Suelos Contaminados, Rendimiento de maíz (kg/ha)		Suelos No Contaminados, Rendimiento de maíz (kg/ha)
1			
2	1,800		2,600
3	1,950		2,750
4	2,050		2,500
5	1,750		2,650
6	1,900		2,600
7	2,100		2,700
8	1,850		2,550
9	1,950		2,800
10	1,900		2,650
11	2,000		2,700

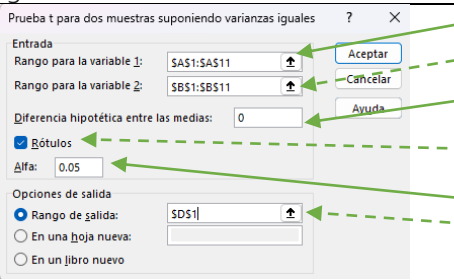


Figura. 5.23. Maíz

C. Plantear Hipótesis
Hipótesis Nula (H_0): No hay un cambio significativo en el rendimiento de maíz entre suelos contaminados y no contaminados. $H_0: \mu_i = \mu_j$
Hipótesis Alternativa (H_a): Hay un cambio significativo en el rendimiento de maíz entre suelos contaminados y no contaminados. $H_a: \mu_i \neq \mu_j$.
D. Regla de Decisión: Si el valor p es menor (<) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

E. Establecer el Nivel de Significancia (α : alfa):
Se utilizará un nivel de significancia del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Prueba t para dos muestras suponiendo varianzas iguales



Suelos Contaminados
Suelos No Contaminados

Diferencia, es cero (0)

Rótulos (✓)

Alfa (α) de 0.05

Salida de datos
Aceptar

Figura. 5.24. Prueba T

G. Resultados del Análisis de Prueba t para dos muestras suponiendo varianzas iguales

D	E	F
Prueba t para dos muestras suponiendo varianzas iguales		
	Suelos Contaminados, Rendimiento de maíz (kg/ha)	Suelos No Contaminados, Rendimiento de maíz (kg/ha)
Media	1925	2650
Varianza	11805.55556	8333.333333
Observaciones	10	10
Varianza agrupada	10069.44444	
Diferencia hipotética de las medias	0	
Grados de libertad	18	
Estadístico t	-16.15549442	
P(T<=t) una cola	1.85036E-12	
Valor crítico de t (una cola)	1.734063607	
P(T<=t) dos colas	3.70071E-12	
Valor crítico de t (dos colas)	2.10092204	

$3.70071\text{E-}12 = 3.70071 \times 10^{-12}$
p valor

H. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor ($<$) que (α) 0.05, se rechazará la (H_0).
Valor $p = 3.70071 \times 10^{-12} < (\alpha) 0.05$, por lo tanto, se rechaza la (H_0).

I. Resultado del análisis: Sí, existe un cambio significativo en el rendimiento de cultivos. Los resultados de la prueba t ($p < 0.05$) indican una diferencia estadísticamente significativa entre los rendimientos medios de maíz en suelos contaminados y no contaminados en el Activo Samaria-Luna, Cunduacán, Tabasco.

5.1.5.6. Prueba t para dos muestras suponiendo varianzas desiguales

La contaminación del suelo por hidrocarburos (TPH) es un problema crítico. Según López-Aguilar *et al.* (2022), la remediación requiere análisis comparativos rigurosos. Se empleará la prueba t para dos muestras suponiendo varianzas desiguales, analizando grupos independientes: suelos tratados y suelos control (no tratados) en ubicaciones distintas. La pregunta de investigación es: ¿Existe una diferencia significativa en la concentración promedio de hidrocarburos totales (TPH) entre suelos tratados y suelos no tratados? Una investigación previa indica que un tratamiento similar reduce el TPH en promedio 200 mg/kg.

75

B. Datos: datos de TPH en (mg/kg) (Cuadro).

	A	B	C
	ID	Suelo Tratado (TPH) (mg/kg)	Suelo Control (TPH) (mg/kg)
1			
2	1	231	764
3	2	352	850
4	3	478	1002
5	4	560	1190
6	5	640	1345
7	6	720	1500
8	7	890	1700
9	8	1020	1920
10	9	1110	2100
11	10	1250	2652



Figura 5.25. Area de hidrocarburos

Cuadro.

C. Plantear Hipótesis

Hipótesis Nula (H_0): No hay diferencia significativa en la concentración promedio de hidrocarburos totales (TPH) entre suelos tratados y suelos no tratados. $H_0: \mu_i = \mu_j$

Hipótesis Alternativa (H_a): Existe una diferencia significativa en la concentración promedio de hidrocarburos totales (TPH) entre suelos tratados y suelos no tratados. $H_a: \mu_i \neq \mu_j$.

D. Regla de Decisión: Si el valor p es menor ($<$) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

Menor ($<$) ; Mayor ($>$)

E. Establecer el Nivel de Significancia (α : alfa):
Se utilizará un nivel de significancia del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Prueba t para dos muestras suponiendo varianzas desiguales

Suelo Tratado
Suelos No Tratado

La Diferencia, es 200 en estudio previo

Rótulos (✓)

Alfa (α) de 0.05

Salida de datos
Aceptar

Figura. 5.26. Prueba t

G. Resultados de la Prueba t para dos muestras suponiendo varianzas desiguales

E	F	G
Prueba t para dos muestras suponiendo varianzas desiguales		
	Suelo Tratado (TPH) (mg/kg)	Suelo Control (TPH) (mg/kg)
Media	725.1	1502.3
Varianza	112972.1	359530.6778
Observaciones	10	10
Diferencia hipotética de las medias	200	
Grados de libertad	14	
Estadístico t	-4.495533883	
P(T<=t) una cola	0.000251739	
Valor crítico de t (una cola)	1.761310136	
P(T<=t) dos colas	0.000503478	
Valor crítico de t (dos colas)	2.144786688	

Cuadro

H. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor ($<$) que (α) 0.05, se rechazará la (H_0).

Valor $p = 0.000503478 < (\alpha) 0.05$, por lo tanto, se rechaza la (H_0).
I. Resultado del análisis: Sí, existe una diferencia altamente significativa ($p=0.0005$; $t=-4.496$); ($gl=14$) en la concentración promedio de TPH entre los suelos tratados (725.1) y control (1502.3), lo que demuestra un efecto evidente del tratamiento.

5.1.5.7. Análisis de varianza de dos factores con varias muestras por grupo.

La relación entre la presencia de aeroalérgenos fúngicos y las alergias respiratorias es un tema de creciente interés en la investigación científica, especialmente en áreas urbanas con alta contaminación ambiental (Cid *et al.*, 2022). El siguiente ejercicio es un supuesto y se apoya en la investigación del efecto de la temporada y el hongo en las alergias respiratorias. Objetivo del estudio: Determinar si la temporada del año y el género de hongo aeroalérgeno principal afectan la incidencia (número de casos registrados) de alergias respiratorias en una clínica de Villahermosa.

Variable Dependiente: Número promedio de pacientes con síntomas de alergia respiratoria por semana.

Factor A: Temporada del año (Niveles: Secas y Lluvias).

Factor B: Género de hongo predominante (Niveles: *Alternaria*, *Aspergillus*, *Cladosporium*).

Réplicas: Se registran datos de 4 semanas diferentes para cada combinación (celda) de temporada y hongo.

1. Hipótesis a Contrastar: El análisis de varianza de dos factores genera tres pruebas de hipótesis simultáneas:

- Hipótesis 1 (Factor A): ¿Existen diferencias significativas en el número de pacientes alérgicos entre la temporada de secas y la de lluvias?
- Hipótesis 2 (Factor B): ¿Alguno de los tres géneros de hongos (*Alternaria*, *Aspergillus*, *Cladosporium*) causa significativamente más casos de alergia que los otros?
- Hipótesis 3 (Interacción A x B): ¿Existe un efecto de interacción entre la temporada del año y el género de hongo sobre el número de pacientes? Es decir, ¿la influencia de un hongo específico en las alergias depende de la temporada en la que se mida? (Por ejemplo: que *Cladosporium* aumente los casos drásticamente en lluvias, pero *Alternaria* los mantenga constantes todo el año).

B. Datos: datos de pacientes con síntomas (Cuadro).

	A	B	C	D		
1	ID	Temporada	Hongo	Pacientes con Síntomas		
2	1	Secas	Alternaria	12		
3	2	Secas	Alternaria	10		
4	3	Secas	Aspergillus	8		
5	4	Secas	Cladosporium	15		
6	5	Lluvias	Alternaria	20		
7	6	Lluvias	Alternaria	18		
8	7	Lluvias	Aspergillus	25		
9	8	Lluvias	Cladosporium	30		
10	9	Secas	Cladosporium	14		
11	10	Secas	Aspergillus	11		
12	11	Lluvias	Cladosporium	28		
13	12	Lluvias	Aspergillus	22		

C. Plantear Hipótesis

Hipótesis 1 (Factor A - Temporada):

Hipótesis Nula (H_0): No hay diferencias significativas entre las temporadas. $H_{01}: \mu_i = \mu_j$

Hipótesis Alternativa (H_a): Existe una diferencia significativa entre las temporadas. $H_{a1}: \mu_i \neq \mu_j$.

Hipótesis 2 (Factor B - Hongo):

Hipótesis Nula (H_{02}): No hay diferencias significativas entre los géneros de hongos. $H_{02}: \mu_i = \mu_j$

Hipótesis Alternativa (H_{a2}): Existe al menos una diferencia significativa entre los géneros de hongos. $H_{a2}: \mu_i \neq \mu_j$

Hipótesis 3 (Interacción A x B):

Hipótesis Nula (H_{03}): No existe interacción entre la temporada y el género de hongo sobre el número de pacientes. $H_{03}: \mu_i = \mu_j$

Hipótesis Alternativa (H_{a3}): Existe interacción entre la temporada y el género de hongo. $H_{a3}: \mu_i \neq \mu_j$

D. Regla de Decisión: Si el valor p es menor (<) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

Menor (<) ; Mayor (>)

E. Establecer el Nivel de Significancia (α : alfa):

Se utilizará un nivel de significancia del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Análisis de varianza de dos factores con varias muestras por grupo.

Análisis de varianza de dos factores con varias muestras por grupo... ? X

Entrada

Rango de entrada:

Fila por muestra:

Alfa:

Aceptar

Cancelar

Ayuda

Opciones de salida

☒ Rango de salida:

☐ En una hoja nueva:

☐ En un libro nuevo

Todo el cuadro

Fila de muestra 2:
Dos réplicas por factor B

Alfa (α) de 0.05

Salida de datos
Aceptar

G. Resultados del Análisis de varianza de dos factores con varias muestras por grupo.

E	F	G	H	I
Análisis de varianza de dos factores con varias muestras por grupo				
RESUMEN	Secas	Lluvias	Total	
Alternaria				
Cuenta	2	2	4	
Suma	22	38	60	
Promedio	11	19	15	
Varianza	2	2	22.6666667	
Aspergillus				
Cuenta	2	2	4	
Suma	19	47	66	
Promedio	9.5	23.5	16.5	
Varianza	4.5	4.5	68.3333333	
Cladosporium				
Cuenta	2	2	4	
Suma	29	58	87	
Promedio	14.5	29	21.75	
Varianza	0.5	2	70.9166667	
Total				
Cuenta	6	6		
Suma	70	143		
Promedio	11.6666667	23.8333333		
Varianza	6.6666667	21.7666667		

ANÁLISIS DE VARIANZA						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Muestra	100.5	2	50.25	19.4516129	0.00238573	5.14325285
Columnas	444.083333	1	444.083333	171.903226	1.2143E-05	5.98737761
Interacción	26.1666667	2	13.0833333	5.06451613	0.05147885	5.14325285
Dentro del grupo	15.5	6	2.58333333			
				Muestra	Columnas	Interacción
Total	586.25	11		H1	H2	H3

H. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor ($<$) que (α) 0.05, se rechazará la (H_0).

H1: Valor $p = 0.002 < (\alpha) 0.05$, por lo tanto, se rechaza la (H_0).

H2: Valor $p = 1.21 \times 10^{-5} < (\alpha) 0.05$, por lo tanto, se rechaza la (H_0).

H3: Valor $p = 0.051 > (\alpha) 0.05$, por lo tanto, se acepta la (H_0).

I. Resultado del análisis: El análisis de varianza de dos factores mostró diferencias significativas en la incidencia de alergias respiratorias entre las temporadas secas y de lluvias, con un valor p de 0.002. También se encontraron diferencias significativas entre los géneros de hongos, con un valor p de 1.21E-05. No obstante, la interacción entre la temporada y el hongo no fue significativa ($p = 0.051$), indicando efectos independientes sobre las alergias.

5.1.5.8. Análisis de varianza de dos factores con una sola muestra por grupo

La optimización de dietas reduce costos significativos en acuicultura, donde la alimentación puede representar hasta el 40% de los gastos de producción (Álvarez-González, 2003). Este ejercicio es un supuesto y evalúa la variabilidad en la actividad enzimática digestiva de larvas de tenguayaca (*Petenia splendida* Günther) alimentadas con distintas dietas en Villahermosa, Tabasco, entre noviembre de 2023 y enero de 2024. Objetivo del estudio: Determinar si la dieta y las condiciones alimenticias afectan la actividad enzimática digestiva en larvas de tenguayaca.

- Variable dependiente: Actividad enzimática digestiva (unidades de actividad por ensayo, UAE).
- Factor A (filas): Tipo de dieta (Niveles: Dieta A —alta proteína, Dieta B —balanceada, Dieta C —alta lípidos).
- Factor B (columnas): Condiciones alimenticias (Niveles: Frecuencia —2/día vs. 4 raciones/día; Tamaño de partícula —fina vs. media). Seleccionar uno para simplificar.

Muestra: Una sola observación por combinación de factores (una medición por celda), con selección aleatoria de larvas.

Medición: Ensayos de digestión estandarizados para cuantificar actividad enzimática, medidas en Unidades de Actividad Enzimática por ensayo (UAE).

1. Hipótesis 1 (Factor A): ¿Existen diferencias significativas en la actividad enzimática entre las dietas A, B y C?
2. Hipótesis 2 (Factor B): ¿Las condiciones alimenticias seleccionadas (frecuencia de ración) generan diferencias significativas en la actividad enzimática?
3. Hipótesis 3 (Interacción A × B): La hipótesis de interacción (A×B) no se incluye, porque el diseño sin replicación no permite estimar ni probar interacción.

B. Datos: datos de dietas racionadas (Cuadro).

	A	B	C
1	Dieta	2 raciones/día	4 raciones/día
2	Dieta A	58.02	69.26
3	Dieta B	77.58	85.39
4	Dieta C	91.84	101.15

C. Plantear Hipótesis

Hipótesis 1 (Factor A - Tipo de dieta):

Hipótesis Nula (H_0) No existen diferencias significativas en la actividad enzimática entre las dietas. $H_{01}: \mu_i = \mu_j$

Hipótesis Alternativa (H_a): Existen diferencias significativas en la actividad enzimática entre las dietas. $H_{a1}: \mu_i \neq \mu_j$.

Hipótesis 2 (Factor B - Condiciones alimenticias):

Hipótesis Nula (H_{02}): Las condiciones alimenticias seleccionadas no generan diferencias significativas en la actividad enzimática. $H_{02}: \mu_i = \mu_j$

Hipótesis Alternativa (H_{a2}): Las condiciones alimenticias seleccionadas generan diferencias significativas en la actividad enzimática. $H_{a2}: \mu_i \neq \mu_j$

Hipótesis 3 (Interacción A x B):

Este diseño sin replicación no permite probar interacción, por lo que no se incluye en el análisis.

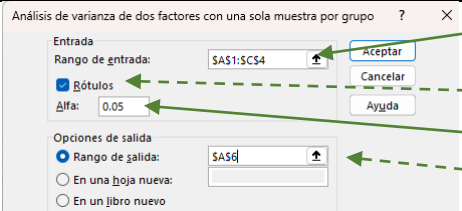
D. Regla de Decisión: Si el valor p es menor (<) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).

Menor (<) ; Mayor (>)

E. Establecer el Nivel de Significancia (α : alfa):

Se utilizará un nivel de significancia del 5% (0.05).

F. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos; Análisis de varianza de dos factores con una sola muestra por grupo



Todo el cuadro
Rótulos
Alfa (α) de 0.05
Salida de datos Aceptar

Figura. 5. 28.

G. Resultados del Análisis de varianza de dos factores con una sola muestra por grupo
H. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor (<) que (α) 0.05, se rechazará la (H ₀).
H1 (resultado en filas): Valor p = 0.0027< (α) 0.05, por lo tanto, se rechaza la (H ₀).
H2 (resultado en columnas): Valor p = 0.0108< (α) 0.05, por lo tanto, se rechaza la (H ₀).
I. Resultado del análisis: Ambos factores, tipo de dieta y frecuencia de alimentación, influyen en la actividad enzimática digestiva de las larvas de tenguayaca (<i>Petenia splendida</i> Günther). Los resultados del ANOVA de dos factores con una sola muestra por grupo, mostraron un p-valor de 0.027 para las dietas y 0.011 para la frecuencia de alimentación, indicando diferencias significativas. Estos hallazgos respaldan la necesidad de realizar ajustes en las dietas para optimizar la producción en acuicultura, confirmando así la pregunta de investigación planteada.

5.1.6. Técnicas Estadísticas para relacionar variables

Supuestos para inferencia: observaciones emparejadas por caso, Normalidad, independencia y sin datos atípicos extremos.

1. ¿Tienes dos variables numéricas medidas en el mismo fenómeno estudiado (pareadas) y ambas varían?

Sí → 2.

No → Esta categoría no aplica (revisa Describir o bien es un tema de asociación no paramétrica).

2. ¿Necesitas una medida sin unidades y acotada entre -1 y 1 para comparar fuerza y dirección entre variables?

Sí → Coeficiente de correlación (Pearson).

No → 3.

3. ¿Buscas cuantificar cuánto varían conjuntamente X y, de forma simultánea, manteniendo las unidades originales de cada variable? (ejemplo: kg y cm: (peso vs longitud de organismo)) para interpretar magnitud con escala propia

Sí → Covarianza.

No → Coeficiente de correlación (Pearson).

5.1.6.1. Coeficiente de Correlación

La moniliasis del cacao es una enfermedad causada por *Moniliophthora roreri*, afectando de manera importante la producción en regiones húmedas del sureste de México. Factores ambientales, como la humedad relativa, se han asociado a la dinámica epidemiológica de la enfermedad, según investigaciones realizadas en cacao (Torres-de la Cruz, M., *et al.* 2020).

El siguiente ejercicio es un **supuesto**, basado en un escenario relacionado con la epidemiología del cacao, el investigador se plantea la siguiente pregunta:

¿En qué medida la variación en la humedad relativa del microambiente del cultivo se asocia linealmente con el incremento de la severidad epidémica de la moniliasis del cacao (*Moniliophthora roreri*) en parcelas evaluadas?

Para contestar la pregunta se usó el **Coeficiente de Correlación** también conocido como **Coeficiente de Correlación de Pearson**, mide la fuerza y dirección de la relación lineal entre dos variables numéricas, un requisito es que ambas sean numéricas, sean de escala de intervalo o de razón y presentar distribución Normal (Stevens S. S., 1946). En este ejemplo hay que determinar si existe una relación lineal entre la **humedad relativa (%)** y la **severidad de la moniliasis del cacao** (escala 1–10) en diferentes parcelas de una plantación.

Variables del estudio

- **Variable independiente:** Humedad relativa (%).

- **Variable dependiente:** Severidad de la moniliasis (1–10).

B. Datos: datos de humedad relativa (%) y severidad de moniliasis (1 poco severo, 10 es muy severo) (Cuadro).

	A	B	C
	Parcela	Humedad relativa (%)	Severidad de moniliasis (1–10)
1			
2	1	65	3
3	2	68	4
4	3	70	4
5	4	72	5
6	5	74	6
7	6	75	6
8	7	78	7
9	8	80	8
10	9	82	9
11	10	85	9

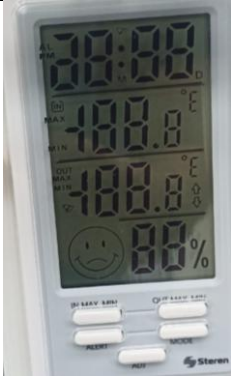


Figura. 5.29.:
Termohigrómetro de medición de humedad relativa (% HR/RH)

C. Plantear Hipótesis
Hipótesis Nula (H_0) No existe correlación lineal entre la humedad relativa y la severidad de la moniliasis. $H_0: r = 0$
Hipótesis Alternativa (H_a): Sí existe correlación lineal entre la humedad relativa y la severidad de la moniliasis. $H_a: r \neq 0$.
D. Regla de Decisión: Si el valor p es menor o igual ($\leq$) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).
Si $p\text{-valor} \leq \alpha$: se Rechaza H_0 (hay correlación significativa)
E. Establecer el Nivel de Significancia (α : alfa): Se utilizará un nivel de significancia del 5% (0.05).

Excel utiliza esta significancia internamente																						
F. Cálculo del valor del p valor Se requiere de un cálculo de un valor t y después este valor se incluye en una segunda ecuación para obtener el valor de p. Para calcular "t", en Excel escribe en una celda vacía: $=r*RAIZ(n-2)/RAIZ(1-r^2)$ $t = 18.548$ Para calcular "p", en Excel escribe en una celda vacía: $= DISTR.T.2C((ABS(t)), (n-2))$ $t = 18.548$ $n = \text{número de pares de datos} = 10$ $p = 7.3583E-08$																						
G. Interpretación del valor del coeficiente de Correlación (r)																						
<table><tr><td>-1.00</td><td>Negativa perfecta</td></tr><tr><td>-0.95</td><td>Negativa muy fuerte</td></tr><tr><td>-0.75</td><td>Negativa considerable</td></tr><tr><td>-0.50</td><td>Negativa media</td></tr><tr><td>-0.10</td><td>Negativa débil</td></tr><tr><td>0.00</td><td>No existe correlación alguna</td></tr><tr><td>0.10</td><td>Positiva débil</td></tr><tr><td>0.50</td><td>Positiva media</td></tr><tr><td>0.75</td><td>Positiva considerable</td></tr><tr><td>0.90</td><td>Positiva muy fuerte</td></tr><tr><td>1.00</td><td>Positivo perfecto</td></tr></table>	-1.00	Negativa perfecta	-0.95	Negativa muy fuerte	-0.75	Negativa considerable	-0.50	Negativa media	-0.10	Negativa débil	0.00	No existe correlación alguna	0.10	Positiva débil	0.50	Positiva media	0.75	Positiva considerable	0.90	Positiva muy fuerte	1.00	Positivo perfecto
-1.00	Negativa perfecta																					
-0.95	Negativa muy fuerte																					
-0.75	Negativa considerable																					
-0.50	Negativa media																					
-0.10	Negativa débil																					
0.00	No existe correlación alguna																					
0.10	Positiva débil																					
0.50	Positiva media																					
0.75	Positiva considerable																					
0.90	Positiva muy fuerte																					
1.00	Positivo perfecto																					
Valor negativo, indica una relación lineal inversa (eso indica que, si una variable sube, la otra baja).																						

Valor positivo, indica una relación lineal directa (Si una variable sube, la otra también).

H. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos (activar si no está); Coeficiente de correlación.

Coeficiente de correlación

Entrada

Rango de entrada:

Agrupado por: ☒ Columnas ☐ Filas

☒ Rótulos en la primera fila

Opciones de salida

☒ Rango de salida:

☐ En una hoja nueva:

☐ En un libro nuevo

Aceptar Cancelar Ayuda

Seleccionamos las dos columnas

Datos en Columnas

Las columnas tienen rótulos o títulos

Salida de datos

Aceptar

Figura. 5.30.

I. Resultados del Análisis de Coeficiente de correlación

E	F	G
	<i>Humedad relativa (%)</i>	<i>Severidad de moniliasis (1-10)</i>
<i>Humedad relativa (%)</i>	1	
<i>Severidad de moniliasis (1-10)</i>	0.988572746	1

La matriz muestra una correlación positiva muy fuerte (0.98) entre humedad y moniliasis; al aumentar una, se incrementa drásticamente la otra.

J. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor o igual ($\leq$) que (α) 0.05, se rechazará la (H_0).

Valor p 7.3583E-08= < (α) 0.05, por lo tanto, se rechaza la (H_0), confirmando una correlación estadísticamente significativa entre las variables.

K. Resultado del análisis: Los resultados demuestran una correlación lineal positiva muy fuerte ($r = 0.98$; $p < 0.05$) entre las variables. Esto indica que el incremento en la humedad relativa del microambiente se correlaciona directamente con el aumento en la severidad de la moniliasis (*Moniliophthora roreri*) en las parcelas.

5.1.6.2. Covarianza

Los derrames de petróleo alteran el carbono orgánico y la materia orgánica del suelo, afectando su estabilidad y dinámica biogeoquímica. Además, generan cambios persistentes en la estructura sedimentaria (López, L., 2021).

El siguiente ejercicio está basado en datos bajo supuestos, donde hay un derrame de petróleo crudo

La pregunta de investigación es:

¿Cómo varían conjuntamente la contaminación por petróleo crudo (mg/kg) y la materia orgánica del suelo (%), en muestras expuestas a distintos niveles del contaminante? (Cov < 0 sugiere relación inversa)

Variables

- Independiente: Contaminación (mg/kg)
- Dependiente: Materia orgánica (%)

A. Datos mínimos

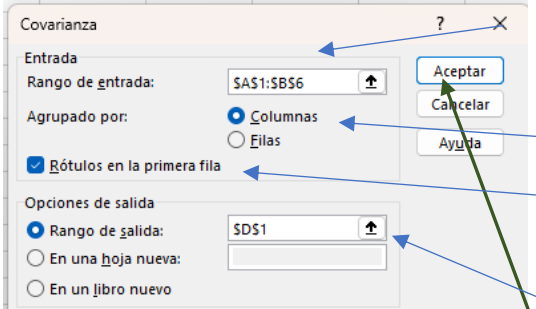
La covarianza requiere mínimo dos pares de datos numéricos emparejados, medidos en las mismas muestras.

B. Datos: datos de contaminación de suelo por hidrocarburo y contenido de materia orgánica (%) (Cuadro).

	A	B	Fig. :
	contaminación por petróleo crudo (mg/kg)	Materia Organica (%)	
1			
2	400	3.9	
3	200	4.7	
4	100	5	
5	500	3.5	
6	300	4.3	

C. Interpretación del valor de Covarianza
El signo negativo (–) de la covarianza indica una asociación lineal inversa entre las variables, lo que significa que, cuando una variable aumenta, la otra tiende a disminuir.
En contraste, un valor positivo de la covarianza refleja una asociación lineal directa, donde ambas variables tienden a aumentar o disminuir de manera conjunta.

D. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos (activar si no está); Covarianza



Seleccionamos las dos columnas

Datos en Columnas

Las columnas tienen rótulos o títulos

Salida de datos

Aceptar

Figura 5.31. Covarianza

E. Resultados del Análisis de Covarianza

D	E	F
	contaminación por petróleo crudo (mg/kg)	Materia Organica (%)
contaminación por petróleo crudo (mg/kg)	20000	
Materia Organica (%)	-76	0.2896

La matriz indica una covarianza negativa (–76) entre la contaminación por petróleo crudo (mg/kg) y la materia orgánica del suelo (%), lo que significa que, en general, a mayores niveles de contaminación corresponde un menor contenido de materia orgánica.

F. Resultado del análisis: La covarianza obtenida describe una asociación lineal inversa entre ambas variables, evidenciando que varían conjuntamente en sentido opuesto. Este resultado permite identificar la dirección de la relación, manteniendo las unidades originales de medición.

5.1.7. Técnicas estadísticas para Predecir

Atributos previos para la Regresión (paramétrica): linealidad, independencia, homocedasticidad, residuos normales (para inferencia).

Para la Suavización exponencial (no paramétrica, disponible en Análisis de datos): serie univariable, corto plazo, sin tendencia ni estacionalidad marcada, no exige normalidad.

Chequeos rápidos: gráfico de línea (tendencia/estacionalidad), dispersión $Y \setminus X$ (linealidad), residuos vs ajustados (patrones).

1. ¿Tienes una serie temporal de una sola variable (solo Y, sin variables explicativas X) y buscas un pronóstico de corto plazo?

Sí → 1.1

No → 2.

1.1 ¿La serie muestra una tendencia clara o patrones que se repiten por temporada?

Sí → Regresión (incluye tiempo y variables indicadoras estacionales).

No → Suavización exponencial.

2. ¿Quieres predecir Y con una o más X (transversal o serie con regresores)?

Sí → Regresión.

No → 3.

3. ¿Solo tienes la serie de Y (sin X) y aun así quieres predecir?

Sí → Suavización exponencial.

No → Replantea objetivo (no hay base para predicción).

5.1.7.1. Regresión

Un estudio de modelación en Nacajuca, mostró que los humedales disminuyeron notablemente: 1,796 ha entre 2000–2008 y 3,995 ha entre 2008–2017. (Palomeque de la Cruz, M. A., *et al.*, 2021). La pérdida de cobertura como variable dependiente muestra cómo la acción humana deteriora ecosistemas. Primero se muestra un antecedente real, y después se trabajará con datos simulados en Huimanguillo para el ejercicio:

93

Pregunta: ¿Cómo se relaciona el crecimiento del área urbana en el municipio de Huimanguillo, Tabasco, con la variación en la extensión de humedales registrada entre los años 2000 y 2010?

Se aplicará Regresión en Excel para modelar la relación entre variables, suponiendo normalidad y homocedasticidad.

A. Variables del estudio

Variable independiente (X): Crecimiento urbano (ha).

Variable dependiente (Y): Pérdida de humedales (ha).

B. Datos: Cuadro ; Crecimiento urbano y pérdida de humedales, 2000–2010 Huimanguillo Tabasco, México,

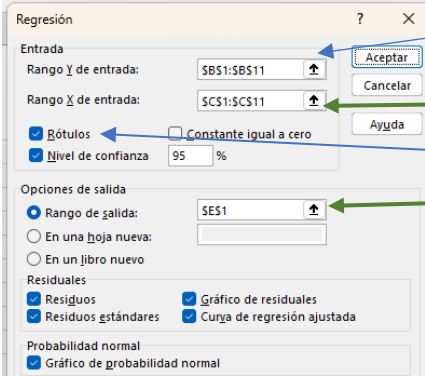
	A	B	C
	Sector	Crecimiento urbano (ha) (X)	Pérdida de humedales (ha) (Y)
1			
2	1	120	290
3	2	150	380
4	3	180	395
5	4	220	430
6	5	260	540
7	6	300	520
8	7	320	590
9	8	350	650
10	9	390	640
11	10	420	710



Figura. 5.32: zona urbana de Huimanguillo, Tabasco México (INEGI, 2010).

C. Plantear Hipótesis
Hipótesis Nula (H_0) El crecimiento urbano no afecta la pérdida de humedales. $H_0: \beta_1 = 0$
Hipótesis Alternativa (H_a): El crecimiento urbano sí afecta la pérdida de humedales $H_a: \beta_1 \neq 0$.
D. Establecer el Nivel de Significancia (α : alfa): Se utilizará un nivel de significancia del 5% (0.05). En Excel colocas un Nivel de Confianza de 95 %
E. Regla de Decisión: Si el valor p es menor o igual ($\leq$) que α (nivel de significancia), se rechazará la hipótesis nula (H_0).
Si $p\text{-valor} \leq \alpha$: se Rechaza H_0 : indicando que la variable independiente (X) sí tiene un efecto lineal sobre la variable dependiente (Y)
F. Cálculo del valor del p valor

El p-valor lo calcula Excel automáticamente y aparece en la columna 'Probabilidad d' junto al coeficiente de la variable analizada (X).		
G. La ecuación del modelo lineal		
La ecuación lineal sirve para predecir Y usando X.		
$Y = \beta_0 + \beta_1 X + \varepsilon$		
Y: variable dependiente (la variable que se quiere predecir)		
X: variable independiente (la variable que se utiliza para predecir)		
β_0 : valor constante del modelo (el valor de Y cuando X es 0)		
β_1 : coeficiente de regresión (valor de la pendiente de la recta)		
ε : error (diferencia entre el valor observado y el valor predicho)		
H. Interpretación del coeficiente de determinación (R^2)		
El valor de R^2 indica el porcentaje de variación de la variable dependiente (X) que es explicado por la regresión.		
Utilidad del modelo	R^2	Interpretación
Poco útil	0.00 – 0.20	El modelo casi no predice Y.
	0.21 – 0.40	La regresión es débil.
Aceptable	0.41 – 0.60	El modelo explica parte importante, pero no es fuerte.
	0.61 – 0.80	El modelo predice bien Y.
Muy buen ajuste	0.81 – 1.00	El modelo predice Y con gran precisión.
Y= variable dependiente; en el ejemplo es la pérdida de humedales (ha)		

I. Cálculo de Valores del Estadístico de Prueba: Ir a Datos; Análisis de datos (activar si no está); Regresión	
	Seleccionamos primero la columna de Y
	Luego la Columna de X
	Las columnas tienen rótulos o títulos
	Salida de datos Selecciona los recuadros marcados con palomita
Acepta	
Figura. 5.33.	
1 Residuos: es para ver la diferencia entre lo real y lo calculado por el modelo. 2 Residuos estándares: detectar datos atípicos (outliers) que se alejan demasiado de la tendencia normal 3 Gráfico de residuales: es vital para confirmar que una línea recta es el modelo adecuado. 4 Curva de regresión ajustada: esto te dará la gráfica con la línea de tendencia ya dibujada. 5 Gráfico de probabilidad normal: para validar estadísticamente tus datos. 6 Intervalos de Confianza: proporcionan un rango de valores para los coeficientes del modelo	
J. Resultados del Análisis de la Regresión	

	E	F	G	H	I	J	K	L	M
2	Resumen								
3									
4	Estadísticas de la regresión								
5	Coefficiente de correlación múltiple	0.9810625							
6	Coefficiente de determinación R ²	0.96248362							
7	R ² ajustado	0.95779407							
8	Error típico	28.1205122							
9	Observaciones	10							
10									
11	ANÁLISIS DE VARIANZA								
12		Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F			
13	Regresión	1	162296.394	162296.394	205.240195	5.50003E-07			
14	Residuos	8	6326.10563	790.763204					
15	Total	9	168622.5						
16									
17		Coefficiente s	Error típico	Estadístico t	Probabilidad d	Inferior 95%	Superior 95%	Inferior 85.0%	Superior 85.0%
18	Intercepción	158.958002	26.3626508	6.02966685	0.00031289	98.16562005	219.750384	98.16562	219.750384
19	Crecimiento urbano (ha) (X)	1.31196389	0.09157784	14.3262066	5.5E-07	1.100784209	1.52314198	1.10078421	1.52314198
20		p(158.958)			p				
21		p(1.311)				el valor de p (5.5E-07) se reporta como p< 0.001 para facilitar su lectura			

	E	F	G	H	I	J	K	L
23	Análisis de los residuales					Resultados de datos de probabilidad		
24								
25	Observación	Pronóstico Pérdida de humedales (ha) (Y)	Residuos	Residuos estándares		Percentil	Pérdida de humedales (ha) (Y)	
26	1	316.393573	-26.393573	-0.99552282		5	290	
27	2	355.752466	24.2475342	0.91457772		15	380	
28	3	395.111359	-0.11135857	-0.00420027		25	395	
29	4	447.589882	-17.5898823	-0.66346187		35	430	
30	5	500.068406	39.931594	1.50615505		45	520	
31	6	552.54693	-32.5469297	-1.22761747		55	540	
32	7	578.786192	11.2138085	0.42296669		65	590	
33	8	618.145084	31.8549157	1.20151582		75	640	
34	9	670.623608	-30.623608	-1.1507289		85	650	
35	10	709.982501	0.0174992	0.00066004		95	710	

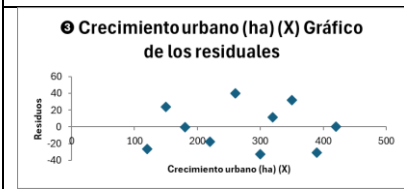


Figura. 5.33.

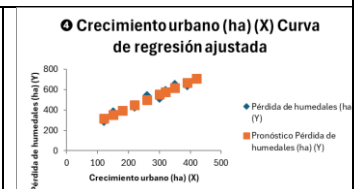


Figura. 5.34.

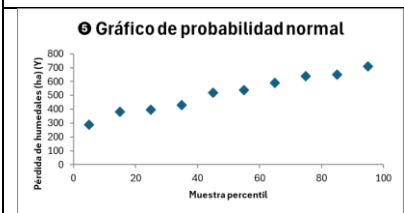


Figura. 5. 35.

K. Contraste de Hipótesis: Según la regla de decisión: Si el p calculado es menor o igual ($\leq$) que (α) 0.05, se rechazará la (H_0). Valor p - 5.5E-07 $\leq$ (α) 0.05, por lo tanto, se rechaza la hipótesis nula (H_0). Los resultados indican una relación lineal significativa entre el crecimiento urbano y la pérdida de humedales.
L. Resultado del análisis: La regresión mostró que el crecimiento urbano aumenta la pérdida de humedales. El modelo es confiable ($R^2=0.95$, $p<0.001$), por lo que para cada hectárea urbana nueva implica perder alrededor de entre 1.1008 y 1.5231 hectáreas de humedales. La ecuación del modelo lineal queda así (ver 4.2.6.): $Y = 158.958 + 1.311X$ Y: variable dependiente, es la pérdida estimada de humedales en ha X: variable independiente, es el crecimiento urbano en ha β_0: valor constante del modelo, 158.958 es el valor de Y cuando X es 0 β_1: coeficiente de regresión, 1.311 es el valor de la pendiente de la recta ϵ: error, solo se coloca en la ecuación teórica, aquí no ➊ Residuos: inicia en G25, muestran la diferencia entre lo real y lo que predijo el modelo, de cada observación. ➋ Residuos estándares: inicia en la H25, valores entre -2 y +2, significa que está dentro de lo normal; valores arriba de ± 2, es un dato atípico. Los resultados indican un buen ajuste del modelo. ➌ Gráfico de residuales: muestra puntos dispersos alrededor de cero sin formar patrones; mientras más cerca estén del cero, mejores son las predicciones del modelo. En el resultado los puntos se dispersan aleatoriamente

alrededor de cero, lo que indican un modelo lineal adecuado.

④ Curva de regresión ajustada: si los puntos observados están cerca de los puntos pronosticados, significa que el modelo lineal predice bien los datos. En la gráfica obtenida, las dos líneas prácticamente coinciden.

⑤ Se interpreta comprobando que los puntos suban de forma continua sin desviarse, indica residuos normales.

El gráfico de probabilidad normal muestra puntos ascendentes sin quiebres, indicando que los residuos se comportan de forma normal y el modelo lineal es válido.

⑥ Intervalos de Confianza (IC): Los intervalos de confianza muestran el rango posible donde se encuentra el valor real del coeficiente: Inferior 95% y Superior 95%.

5.1.7.2. Suavización exponencial

La predicción del caudal hidrológico es esencial para la gestión del agua, pues permite anticipar variaciones críticas del flujo, mejorar la toma de decisiones y reducir riesgos asociados a eventos extremos, elementos clave para la seguridad hídrica regional (Workneh, H. A., & Jha, M. K., 2025).

5.1.8. Técnicas estadísticas para Descubrir patrones especiales

Regresión: requiere linealidad, Normalidad, independencia, homocedasticidad.

No requieren Normalidad.

Suavización exponencial simple, Media móvil y Análisis de Fourier, Muestra y Generación de números aleatorios.

101

1. ¿Tus datos están ordenados en el tiempo (serie temporal)?
Sí → 2.
No → 4.
2. ¿Quieres reducir las oscilaciones de corto plazo para resaltar la tendencia general?
Sí → Media móvil.
No → 3.
3. ¿Quieres detectar ciclos o estacionalidad (patrones repetidos en el tiempo)?
Sí → Análisis de Fourier.
No → Replantea objetivo (quizá Describir o Comparar).
4. ¿Necesitas seleccionar, al azar, una parte representativa de un conjunto real (parcelas, sitios, individuos)?
Sí → Muestra.
No → 5.
5. ¿Quieres crear datos simulados para pruebas, escenarios o capacitación (sin usar datos reales)?
Sí → Generación de números aleatorios.
No → Replantea objetivo (no hay técnica de este grupo que aplique).

Estadística aplicada a las ciencias ambientales y de la vida: una perspectiva computacional con Excel.

5.1.8.1. Media móvil

¿Sabías que el ruido visual puede ocultar un peligro ambiental? La media móvil es tu mejor aliada para suavizar variaciones aleatorias y revelar la tendencia real de contaminantes en ecosistemas. La NOM-127-SSA1-2021, establece un límite permisible de 11.00 mg/L de nitrógeno de nitratos (N-NO₃). A manera de ejemplo se simuló el registro de un contaminante ambiental a lo largo del tiempo (concentración mensual de Nitrato "N-NO₃" en un río "mg/L"). Analizaremos una simulación mensual en un río. El reto: filtrar las oscilaciones de corto plazo para identificar si hay un patrón y ver si el agua que consumimos cumple realmente con la Normatividad.

B. Datos: Concentración mensual de Nitratos en el río (mg/L) (Cuadro).

	A	B
1	Mes	Concentración (mg/L)
2	1	6.1
3	2	7.4
4	3	9.7
5	4	11.2
6	5	11.3
7	6	10.7
8	7	7.5
9	8	10.9
10	9	12.6
11	10	13.8



Figura 5.36.

C. Configuración de entradas:

Entra a Datos: elije, análisis de datos, luego, media móvil

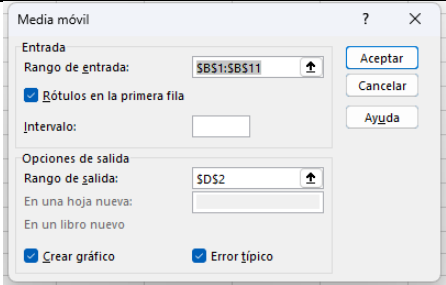


Figura. 5.37.

Rango: columna B, Datos de Concentración

Rótulos: selecciónala

Intervalo: déjalo sin seleccionar

Rango de salida: en la misma hoja, celda "D2"

Crear gráfico: selecciónalo

Error típico: selecciónalo

D. Resultados		El resultado inicia en la celda D2, Los títulos en amarillo escríbelo tu Las celdas con "#N/D", están bien, el procedimiento no permite su calculo La media es sumativa, va cambiando, conformé avanzan los meses
D	E	
Media movil	Error tipico	
#N/D	#N/D	
#N/D	#N/D	
7.733333333	#N/D	
9.433333333	#N/D	
10.73333333	1.5609826	
11.06666667	1.09188929	
9.833333333	1.40237893	
9.7	1.52958479	
10.33333333	2.001851	
12.43333333	1.67785136	

Media móvil

Punto de datos	Real	Pronóstico
1	5.73	5.73
2	7.23	7.23
3	9.43	7.73
4	11.07	9.43
5	10.73	10.73
6	10.73	10.73
7	7.23	9.43
8	10.73	9.43
9	12.43	10.73
10	13.93	12.43

Figura. 5.38.

La media móvil evidenció una tendencia ascendente estadísticamente relevante en nitratos, mostrando una tendencia ascendente que podría representar un posible riesgo sanitario al compararse con los límites normativos.

La media móvil revela, en el gráfico un patrón ascendente sostenido de nitratos, mitigando ruido y evidenciando posible incumplimiento normativo sanitario, al rebasar el límite permisible de 11 mg/L. La media móvil permite identificar que el rebase no es un evento aislado, sino parte de una tendencia persistente.

5.1.8.2. Análisis de Fourier

El análisis de Fourier permite descomponer series temporales en frecuencias, identificando ciclos y estacionalidades ocultas en datos ambientales.

¿Es posible determinar la periodicidad y magnitud de las oscilaciones térmicas diarias en un ecosistema mediante el análisis de frecuencias para predecir comportamientos biológicos? El Análisis de Fourier descompone señales complejas en ondas simples, revelando ciclos dominantes y amplitudes de variación.

B. Datos: tabla de lecturas de temperatura (°C) (Cuadro)

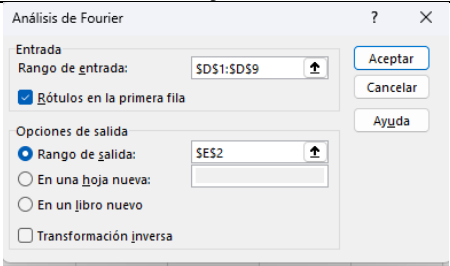
	A	B	C	D
1	Tiempo	Día	Fase del Ciclo	°C
2	1	1	Mañana	25
3	2	1	Tarde	35
4	3	1	Noche	20
5	4	1	Madrugada	15
6	5	2	Mañana	25
7	6	2	Tarde	35
8	7	2	Noche	20
9	8	2	Madrugada	15



Figura. 5.39.

C. Configuración de entradas:

Entra a Datos: elije, análisis de datos; Análisis de Fourier



Rango: columna D,
Datos de Temperatura °C

Rótulos: selecciónala

Rango de salida: en la
misma hoja, celda "E2"

Transformación inversa:
sin seleccionar

Figura 5.40.

D. Resultados	El resultado inicia en la celda E2. Genera datos de frecuencia y amplitud
---------------	---

	A	B	C	D	E
1	Tiempo	Día	Fase del Ciclo	°C	Análisis de Fourier
2	1	1	Mañana	25	190
3	2	1	Tarde	35	0
4	3	1	Noche	20	9.99999999999997-40i
5	4	1	Madrugada	15	0
6	5	2	Mañana	25	-10
7	6	2	Tarde	35	0
8	7	2	Noche	20	10+40i
9	8	2	Madrugada	15	0

para crear dos gráficas que identifiquen ciclos.

Se calcula la columna F "Frecuencias" es para que el eje X de la gráfica muestre "Ciclos por hora": $= (A2-1) / (8 * 6)$
Frecuencia = (Índice) / (Tiempo Total del Estudio "TTT")
TTT= 8 datos totales por cada 6 horas.

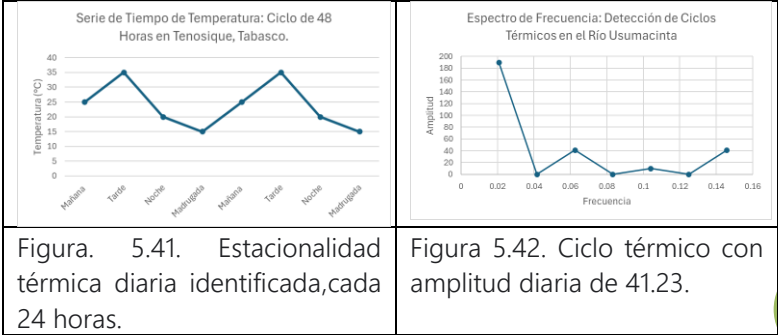
La columna G "Amplitud", se calcula con $=IM.ABS(E2)$

G2	:	x	✓	f _x	$=IM.ABS(E2)$		
	A	B	C	D	E	F	G
1	Tiempo	Día	Fase del Ciclo	°C	Análisis de Fourier	Frecuencia	Amplitud
2	1	1	Mañana	25	190	0	190
3	2	1	Tarde	35	0	0.020833333	0
4	3	1	Noche	20	9.99999999999997-40i	0.041666667	41.231056
5	4	1	Madrugada	15	0	0.0625	0
6	5	2	Mañana	25	-10	0.083333333	10
7	6	2	Tarde	35	0	0.104166667	0
8	7	2	Noche	20	10+40i	0.125	41.231056
9	8	2	Madrugada	15	0	0.145833333	0

La transformada de Fourier confirmó la periodicidad diaria del ecosistema, revelando una magnitud de 41.23.

Selecciona tus datos: Sombrea el rango de la columna "D"
Inserta el gráfico, de serie de Dominio del Tiempo, el cual muestra la evolución de los datos conforme avanza el reloj: Ve a la pestaña Insertar > Gráficos recomendados > Selecciona Líneas con marcadores.
Dentro del gráfico en el ejehorizontal da clic derecho, luego Diseño Gráfico> Datos> Seleccionar datos> Etiquetas del eje horizontal (categoría)> Editar> elije los datos de Fase del Ciclo (mañana, tarde, noche, madrugada).

Segundo Gráfico, serie de Dominio de la Frecuencia, el cual muestra la descomposición de los datos en ciclos o ritmos: Insertar> Gráficos recomendados> Todos los gráficos> Selecciona Dispersion con Líneas rectas y marcadores



5.1.8.3. Muestra

La opción Muestreo en Excel permite seleccionar submuestras estadísticas representativas para aplicar la teoría del muestreo en estudios científicos. En el monitoreo ambiental, el muestreo aleatorio se usa para estimar contaminantes sin analizar todo el ecosistema. En este ejercicio se aplicará muestreo aleatorio con datos simulados en Excel. A. Imagina que registraste edades de mojarra en dos zonas de muestreo ambiental, no quieres analizarlos todos, solo una muestra representativa. ¿Cuáles de los localizados hay que analizar?

B. Datos: Edades de mojarra (*Oreochromis niloticus*) en zonas de muestreo fluvial (años) (Cuadro).

	A	B	C	
	Zona de muestreo	ID del organismo	Edad (años)	
1				
2	Río arriba	M1	2.1	
3	Río arriba	M2	2.4	
4	Río abajo	M3	1.9	
5	Río abajo	M4	2.7	
6	Río abajo	M5	2.2	



Figura 5.43.

C. Configuración de entradas	
Muestra ? × Entrada Rango de entrada: SC\$1:SC\$6 <input checked="" type="button" value="Aceptar"/> Cancelar Ayuda Método de muestreo ☐ Periódico ☒ Aleatorio Período: Número de muestras: 2 Opciones de salida ☒ Rango de salida: ☐ En una hoja nueva: ☐ En un libro nuevo: SES1 Rango: columna de edades Rótulos: selecciónala Método de muestreo: es aleatorio y para este caso son dos ejemplares a analizar Rango de salida: en la misma página, en la celda	
Figura. 5.44.	
D. Resultados	Nos da dos datos de la base de manera aleatoria. Útil si la base es más grande
E 2.4 2.7	

5.1.8.4. Generación de números aleatorios

La opción de Generación de números aleatorios permite crear conjuntos de datos simulados en Excel para el análisis de escenarios, la validación de modelos mediante la evaluación de precisión y confiabilidad de resultados, y la capacitación de usuarios, sin utilizar información real, reduciendo riesgos y optimizando el tiempo, además de facilitar simulaciones estadísticas y apoyar la toma de decisiones informadas.

Generación de números aleatorios: la información que requiere y se puede elegir

Número de variables: genera el número de columnas

Cantidad de números aleatorios: define la cantidad de filas

Distribución: es la forma en que los valores de un grupo se organizan, en este caso hay 7 tipos de las distribuciones:

Uniforme: Todos los resultados tienen exactamente la misma probabilidad de ocurrir en un experimento dado.

Normal: Distribución en forma de campana donde los datos se agrupan cerca del promedio central.

Bernoulli: Experimento aleatorio que solo tiene dos resultados posibles, llamados éxito o fracaso.

Binomial: Indica la probabilidad de obtener un número de éxitos en varios ensayos de Bernoulli.

Poisson: Representa el número de veces que ocurre un evento en un intervalo de tiempo.

Frecuencia relativa: Es el cociente obtenido al dividir la frecuencia absoluta entre el número total.

Discreta: Modelo que describe variables cuyos resultados posibles son números enteros, aislados y finitos.

Parámetros:

Rango de entrada de valores y probabilidades; debes seleccionar un rango de dos columnas.

Primera columna: Los valores numéricos que quieres generar (por ejemplo: 1, 2, 3).

Segunda columna: La probabilidad de que ocurra cada valor

Iniciar con: el punto de partida, con un numero entero

Rango de salida: se coloca la celda donde se iniciará la colocación de los datos

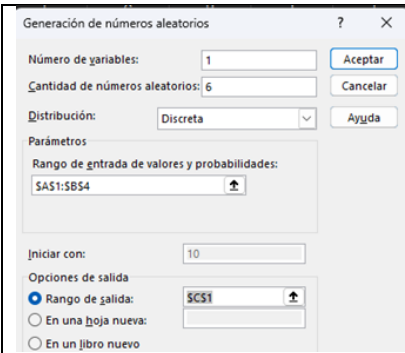


Figura. 5.45.

	A	B
1	Valores A	Probabilidades
2	10	0.15
3	20	0.3
4	30	0.15
5	20	0.1
6	10	0.15
7	15	0.15

	A	B	C
1	Valores A	Probabilidades	
2	10	0.15	10
3	20	0.3	30
4	30	0.15	10
5	20	0.1	20
6	10	0.15	20
7	15	0.15	30

Número de variables: 1
Cantidad de números aleatorios: 6
Distribución: Discreta
Parámetros: Rango de entrada de valores y probabilidades: \$A\$1:\$B\$4
Iniciar con: 10
Rango de salida: \$C\$1
Debe haber dos columnas
Con 6 líneas, como la cantidad solicitada
La suma de probabilidad debe ser igual a 1
Genero 6 datos de distribución Discreta
Puede generar miles de filas

Hay 6 tipos de distribuciones, más

5.2. Estadística No Paramétrica

La estadística no paramétrica es una rama que no exige que los datos sigan una distribución específica (como la normal) ni asume parámetros fijos. Se basa principalmente en el orden y rango de los valores, lo que la hace ideal para muestras pequeñas o datos cualitativos. Es la opción más robusta cuando los datos son irregulares o contienen valores atípicos que sesgarían un análisis tradicional. Para aplicar estas pruebas fácilmente, descarga del internet el archivo de "Real Statistics" en tu computadora, luego en Excel dirígete a Archivo, Opciones, Complementos, selecciona "Ir" y usa "Examinar" para cargar el recurso descargado.

En el encontraras técnicas de análisis no paramétrico fáciles de usar en Excel. En esta versión del libro no lo manejaremos.

Referencias

Álvarez-Coronel, G., Domínguez-Rodríguez, V. I., Adams, R. H., Palma-López, D. J., Zavala-Cruz, J. & Gaspar-Génico J.A. (2023). Fertility impacts in crude oil-contaminated soil based on type and quantity of clay. *International Journal of Environmental Science and Technology*, 20, 9555–9570.. <https://doi.org/10.1007/s13762-022-04616-8>

113

Álvarez-González, C. A. (2003). Actividad enzimática digestiva y evaluación de dietas para el destete de larvas de la cabrilla arenera *Paralabrax maculatofasciatus* (Percoidei: Serranidae) (Tesis doctoral). Centro Interdisciplinario de Ciencias Marinas, Instituto Politécnico Nacional, La Paz, Baja California Sur, México.

Álvarez-Suárez, M. L., García, A., Flores-Magdaleno, H., & López-Mata, L. (2022). Análisis de la fertilidad del suelo mediante la validación e interpolación Kriging de sus variables. *Revista Terra Latinoamericana*, 40(1), 1–15. (estudio con N, P, K en Hidalgo, con métodos y rangos típicos)

Carreño-Ruiz, S. D., Cappello-García, S., Gaitán-Hernández, R., Torres-De la Cruz, M., Gaspar-Génico, J. A., & Rosique-Gil, J. E. (2020). Production of basidiomes of *Schizophyllum commune* (Fungi: Basidiomycota) in agricultural by-products of Tabasco, Mexico. *Agroproductividad*, 13(5), 65-71. <https://doi.org/10.32854/agrop.vi.1606>

Cochran, W. G. (1977). Sampling techniques (3ª ed.). John Wiley & Sons.

CONAGUA, 2022. Estadísticas del Agua en México.

Cid Martínez, M. A., Hernández González, J. R., & Gaspar Génico, J. Á. (2022). Aeroalérgenos (*Curvularia*, *Nigrospora* & *Torula*) en una zona urbana de Villahermosa, Tabasco, México. *Kuxulkab'*, 28(62), e5062.
<https://doi.org/10.19136/kuxulkab.a28n62.5062>

CONAPESCA. (2006). Anuario estadístico de pesca 2006: Especies objetivo y asociadas en pesquerías del Golfo de México (Tabasco, Campeche, Yucatán). Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación. Información de capturas en Tabasco: bagre marinus (*Ariopsis felis*), *Lutjanus griseus*, *Caranx hippos*, entre otras

Corella Justavino, F., Valdez Hernández, J. I., Cetina Alcalá, V. M., González Cossío, F. V., Trinidad Santos, A., & Aguirre Rivera, J. R. (2001). Estructura forestal de un bosque de mangles en el noreste del estado de Tabasco, México. *Ciencia Forestal en México*, 26(90), 73–102.

Freedman, D., Pisani, R., y Purves, R. (2007). Estadística (4.^a ed.). W. W. Norton & Company.

Harshaw, L. T., Larkin, I. V., Bonde, R. K., Deutsch, C. J., & Hill, R. C. (2016). Morphometric body condition indices of wild Florida manatees (*Trichechus manatus latirostris*). *Aquatic Mammals*, 42(4), 428–439.
<https://doi.org/10.1578/AM.42.4.2016.428>

Instituto Nacional de Estadística y Geografía. (2010). *Cartografía geoestadística urbana: Cierre del Censo de Población y Vivienda 2010. Huimanguillo* [Conjunto de datos vectoriales]. www.inegi.org.mx

Investing.com. (2025). *US Cocoa Futures Historical Data* [Datos de apertura, máximo, mínimo y cierre para julio 2025]. Recuperado de <https://www.investing.com/commodities/us-cocoa-historical-data>

IQAir. (2025). *Air quality index (AQI⁺) in Tabasco, Mexico* [<https://www.iqair.com/>]. Recuperado el 3 de agosto de 2025, de IQAir: Tabasco, México

Kish, L. (1965). *Survey sampling*. John Wiley & Sons.

López-Aguilar, S., Domínguez-Rodríguez, V. I., Gaspar Génico, J. A., Zavala-Cruz, J., Hernández-Nataren, E., & Adams, R. H. (2022). The potential of the osmological perception of landholders and managers as a practical way of assessing the impact of oil spills on soil. *Soil Security*,

López-Luna, M. A., & García-Grajales, J. (2010). Análisis bibliográfico del conocimiento de los crocodilianos en México. *Revista Latinoamericana de Conservación*, 1(2), 45–60.

López, L. (2021). Derrames de petróleo y sus efectos en el carbono orgánico de suelos y sedimentos. *Boletín de la Academia de Ciencias Físicas, Matemáticas y Naturales*, 81(1), 25–28.

Magaña-Alejandro, M. A., Gama Campillo, L. M., & Mariaca Méndez, R. (2010). *El uso de las plantas medicinales en las comunidades Maya-Chontales de Nacajuca, Tabasco, México*. *Polibotánica*, 29, 213–262.

Mendenhall, Beaver y Beaver (2010). Introducción a la probabilidad y estadística. CENGAGE Learning.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4), 558–625.

Palomeque de la Cruz, M. A., Ruiz Acosta, S. del C., Ramos Reyes, R., Magaña Alejandro, M., & Galindo-Alcántara, A. (2021). *Modelación de cambios de coberturas y uso de suelo en Nacajuca, Tabasco*. *Revista Mexicana de Ciencias Agrícolas*, 12(4). <https://doi.org/10.29312/remexca.v12i4.2622>

Ramírez-Guillermo, M. Á., de la Rosa-Santamaría, R., & Gutiérrez, O. A. (2018). *Variación morfológica de frutos y semillas de cacao (Theobroma cacao L.) de plantaciones en Tabasco, México*. *Revista Fitotecnia Mexicana*, 41(2), 117–125.

Ruiz-Álvarez, O., Arteaga-Ramírez, R., Vázquez-Peña, M. A., Ontiveros Capurata, R. E., & López-López, R. (2012). Balance hídrico y clasificación climática del estado de Tabasco, México. *Universidad y Ciencia*, 28(1), 1–14.
<https://www.scielo.org.mx/>

Servicio Meteorológico Nacional (SMN–CONAGUA), 2025. *Normales climatológicas por estado: Tabasco* [Base de datos]. CONAGUA.
<https://wwwsmn.cna.gob.mx/es/climatologia/informacion-climatologica/normales-climatologicas-por-estado>

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677-680.

<https://www.jstor.org/stable/1671815>

Torres-de la Cruz, M., Mora-Aguilera, G., Ortiz-García, C.F., Cruz-Pérez, Aracely De la Cruz Pérez A., & Gaspar-Génico, J.Á. (2020). Flujos productivos determinan la estructura epidémica de la moniliasis del cacao en el sureste de México. *Revista fitotecnia mexicana*, 43(4), 421-431. Epub 14 de agosto de 2023. <https://doi.org/10.35196/rfm.2020.4.431>.

Wackerly, Mendenhall y Scheaffer (2010). *Estadística matemática con aplicaciones*. CENGAGE Learning

Workneh, H. A., & Jha, M. K. (2025). *Utilizing deep learning models to predict streamflow*. *Water*, 17(5), 756.

<https://doi.org/10.3390/w17050756>

Estadística aplicada a las ciencias ambientales y de la vida: una perspectiva computacional con Excel.

Índice Analítico

A Aleatoriedad, 7 Análisis de Fourier, 114–116 Análisis de varianza (ANOVA), 72–77 Asimetría, 33 B Bernoulli (distribución), 119 Binomial (distribución), 119 Box plot (gráfica de cajas y bigotes), 51 Bondad de ajuste (R^2), 104 C Coeficiente de correlación (Pearson), 94–98 Coeficiente de determinación, 104–108 Covarianza, 94, 99–100 Curtosis, 33 D Desviación estándar, 33, 56 Distribución normal, 63 Distribuciones de probabilidad, 119 Datos cualitativos y cuantitativos, 18–19	E Error tipo I, 59–62 Error tipo II, 60–62 Estimación de parámetros, 47, 56 Escalas de medición, 20–21 F Frecuencia relativa, 119 Función de densidad, 63 Función de distribución, 63 Factor (en ANOVA), 76–93 G Gráfica de dispersión, 34, 43 Gráfica de histograma, 41, 50, 67–68 Gráfica de cajas y bigotes, 42, 51 Gráfica de cascada, 43, 52 H Hipótesis alternativa (H_a), 48, 57 Hipótesis nula (H_0), 48, 57 Homocedasticidad, 54, 63 Histograma, 41, 50
--	--

I Inferencia estadística, 5, 14, 47 Intervalos de confianza, 47, 56–57 Incertidumbre, 47, 56 Independencia estadística, 94, 101	O Ordenamiento de datos, 13, 22 Ordinal (escala), 20 Observaciones emparejadas, 72, 78
J	P Percentil, 33, 69–71 Población, 7, 16 Probabilidad, 5, 63 Pruebas de hipótesis, 48–62
K	
L Linealidad, 101–105 Media, 33 Media móvil, 110–113	Q Cuartiles, 33
M Matriz de correlación, 98 Mediana, 33 Moda, 33 Muestreo aleatorio, 6–7, 16	R Rango intercuartílico, 33 Regresión lineal, 101–108 Residuales, 105–108 Razón (escala), 20
N Nivel de significancia (α), 48, 57–59 Normalidad, 54, 63 Números aleatorios, 119–120 No paramétrica (estadística), 121	S Serie temporal, 101, 110 Suavización exponencial, 101, 109 Sesgo estadístico, 7 Supuestos estadísticos, 54, 94

T	W
Teoría del muestreo, 6–7, 15–17	X
T de Student, 78–85	Y
Tendencia central, 33	Z
Transformada de Fourier, 114–116	Prueba Z, 80–81
U	
V	
Valor p, 50, 59–62	
Varianza, 33, 56	
Variables cualitativas, 18	
Variables cuantitativas, 18	

Estadística aplicada a las ciencias ambientales y de la vida: una perspectiva computacional con Excel.

Colofón

Esta obra se terminó de elaborar y editar en el año 2026, en Villahermosa, Centro, Tabasco, México.

El texto fue dictaminado mediante el sistema académico de pares ciegos y se publica en formato impreso y digital.

La impresión y composición editorial se realizaron utilizando tipografía académica estándar y materiales de uso editorial universitario.

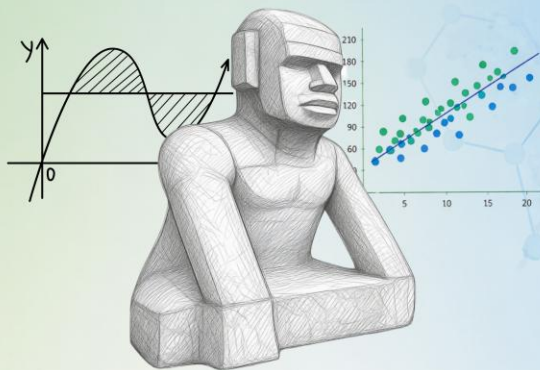
El diseño de la portada estuvo a cargo de Stephanie Pérez Alamilla y la fotografía del autor es de José Ángel Gaspar Genico.

Obra registrada ante el Instituto Nacional del Derecho de Autor (INDAUTOR).

Todos los derechos reservados conforme a la Ley Federal del Derecho de Autor.

ISBN versión impresa: 978-970-96059-0-7

ISBN versión digital:



Dirigido a universitarios que desean aplicar la estadística en su formación profesional. Contiene ejercicios prácticos que guían al lector paso a paso e invita a hacer anotaciones en la edición impresa. Revisado por profesionales en estadística para asegurar claridad y confianza.

Este libro te permitirá:

- Comprender la aplicación de la estadística descriptiva
- Comprender los fundamentos de la estadística aplicada
- Analizar datos socioeconómicos, ambientales y biológicos con Excel®
- Interpretar resultados de manera clara y objetiva
- Aplicar herramientas estadísticas en contextos reales.

ISBN 978-970-96059-0-7



9 789709 605907 >



ASOCIACIÓN MEXICANA DE ESTADÍSTICA
APLICADA E INTELIGENCIA ARTIFICIAL, A.C.

AMEAIA